

# *Variability of urban surface temperatures and implications for aerodynamic energy exchange in unstable conditions*

Article

Accepted Version

Crawford, B., Grimmond, S. B., Gabey, A., Marconcini, M., Ward, H. C. and Kent, C. W. (2018) Variability of urban surface temperatures and implications for aerodynamic energy exchange in unstable conditions. Quarterly Journal of the Royal Meteorological Society, 144 (715). pp. 1719-1741. ISSN 1477-870X doi: <https://doi.org/10.1002/qj.3325> Available at <https://centaur.reading.ac.uk/77578/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/qj.3325>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

*Variability of urban surface temperatures and implications for aerodynamic energy exchange in unstable conditions*

B. Crawford<sup>\*1,2</sup>, C.S.B. Grimmond<sup>1</sup>, A. Gabey<sup>1,3</sup>, M. Marconcini<sup>4</sup>, H.C. Ward<sup>1,5</sup>, C. W. Kent<sup>1</sup>

\*Corresponding author

<sup>1</sup>Department of Meteorology, University of Reading, UK

<sup>2</sup>Current affiliation: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, USA

<sup>3</sup>Current affiliation: Institute for Environmental Analytics, Reading, UK

<sup>4</sup>Remote Sensing Data Center, German Aerospace Center (DLR), Germany

<sup>5</sup>Current affiliation: Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria

## Abstract

Sensible heat flux ( $Q_H$ ) is a critical driver of surface and boundary layer meteorological processes, especially in urban areas. Aerodynamic resistance methods (ARM) to model  $Q_H$  are promising because, in principle, all that is needed is surface temperature ( $T_o$ ), air temperature ( $T_A$ ) and an aerodynamic resistance term ( $r_H$ ). There are significant challenges in urban areas however, due to uncertainties in satellite-derived land surface temperatures ( $LST$ ), logistical challenges to obtain high-resolution air temperatures, and limited understanding of spatial and temporal variability of  $r_H$  and associated variables (e.g. thermal roughness length). This work uses an extensive  $LST$  dataset covering six years (2011-2016) in central London and a long-term *in situ* observation network to analyse variability of  $LST$  and  $r_H$  variables. Results show that  $LST$  is spatially correlated with building and vegetation land cover with coherent thermal structures at length scales less than 500-1000 m. Additionally, satellite-observed  $LST$  varies with average building height (up to 10% cooler in areas with tall buildings). The  $r_H$  term and associated variables are observed to vary on daily and seasonal cycles and findings are used to model  $Q_H$  using five variations of an ARM-based approach on a 100 m pixel basis. Modelled  $Q_H$  is compared to observations from three scintillometer paths and an eddy covariance flux tower. We find generally good agreement between observations and models, though there is uncertainty in all methods (mean absolute error ranges from 58.1-129.3  $W m^{-2}$ ) due to challenges in determining high-resolution meteorological and surface inputs, particularly  $LST$  and friction velocity ( $u_*$ ). Additional complexity in evaluating modelled  $Q_H$  arises from anthropogenic heat sources: long-term tower-based observations show that  $T_A$  and radiometer-derived  $T_o$  are warmer during working weekdays than non-working days (up to 0.7°C) and that there is an observed lag (2-3 hours) between energy consumption and observed warming and  $Q_H$ .

**Keywords:** urban meteorology, surface energy balance, remote sensing, observations

## 1. Introduction

Turbulent sensible heat flux ( $Q_H$ ,  $W m^{-2}$ ) is a dominant component of surface-atmosphere energy exchange in many cities due to relative lack of moisture and vegetation compared to non-urban environments. Knowledge of  $Q_H$  is important for modelling dynamic processes in the urban atmosphere and partitioning net energy exchange between additional surface energy balance processes.  $Q_H$  is also a fundamental driver of urban boundary layer growth and hence a primary control on variations in urban air quality.

Recent progress in modelling urban surfaces has highlighted uncertainties in determining  $Q_H$  using both multi-layer canopy and bulk models (Grimmond et al. 2011). Additionally, characterization of sub-kilometre variability of urban physical characteristics and its impact on energy fluxes is an ongoing challenge to parameterise urban turbulent flux terms for meso-scale models (Barlow et al., 2017). Alternative techniques that can calculate high-resolution (<1 km)  $Q_H$  are needed to compare with urban land surface models.

An aerodynamic resistance method (ARM) is a common approach for modelling surface fluxes and this technique continues to be used in current meso-scale numerical weather prediction models (Mahrt 1996, Chen et al. 2010). However, high-resolution (e.g. sub-city scale) application of this technique has been limited in cities due to the extreme heterogeneity of the urban surface at numerous length scales.

The bulk aerodynamic resistance approach is based on surface layer similarity theory where  $Q_H$  is calculated as:

$$Q_H = c_p \rho \frac{T_o - T_A}{r_H} \quad (1)$$

where  $c_p$  is heat capacity of air ( $J kg^{-1} K^{-1}$ ),  $\rho$  is air density ( $kg m^{-3}$ ),  $T_o$  is the aerodynamic surface temperature (K),  $T_A$  is air temperature (K), and  $r_H$  is the bulk aerodynamic resistance for heat ( $s m^{-1}$ ) of the complete 3D urban surface.

The aerodynamic surface temperature ( $T_o$ ) is the air temperature extrapolated logarithmically to the height level  $z_d + z_H$ , where  $z_d$  is displacement height and  $z_H$  is the roughness length for heat (e.g. Kanda et al. 2007). Practically,  $T_o$  is difficult to measure and is often replaced in eq. 1 with some other surface temperature, such as satellite-observed land surface temperature ( $LST$ ), the radiative surface temperature ( $T_R$ ) calculated from measurements of upwelling longwave radiation, or area-weighted facet-based complete surface temperatures from some combination of thermal camera measurements (Voogt and Grimmond 2000) and 3D surface models (Morrison et al. 2018).

Large diurnal variations in  $z_H$  have been observed previously in urban areas (Voogt and Grimmond 2000, Moriwaki and Kanda 2006, Kanda et al. 2007), as well as in bare soil and vegetated environments (Verhoef et al. 1997). Brutsaert and Sugita (1996) attribute this variation to changes in solar angle which leads to vertical movement of the heat source in the canopy through the day. A relation between  $z_H$  and solar angle was observed in a light industrial area of Vancouver, Canada (Voogt and Grimmond 2000), but not over a 1:5 scale urban model for a full year (Kanda et al. 2007). Other explanations for  $z_H$  variation include use of inappropriate surface temperatures (Kanda et al. 2007) or anisotropic effects of canopy vegetation (e.g. Qualls and Hopson 1998). Overall, these large diurnal variations are difficult to interpret and there is significant uncertainty in  $z_H$  derived from observations due to accumulated measurement and source area uncertainties. In a traditional Monin-Obukhov similarity theory (MOST) framework, the aerodynamic resistance term  $r_H$  (eq. 1) can be defined as (Verma, 1989):

$$r_H = \frac{1}{ku^*} \left[ \ln \left( \frac{z'}{z_0} \right) - \Psi_h \left( \frac{z'}{L} \right) \right] + \frac{1}{ku^*} \ln \left( \frac{z_0}{z_H} \right) \quad (2)$$

where  $z_0$  is the roughness length for momentum,  $z' = z_m - z_d$  ( $z_m$  is measurement height),  $\Psi_h$  is a similarity theory stability function,  $L$  is the Obukhov Length,  $k$  is von Karman's constant (0.4), and  $u^*$  is friction velocity. For convenience, the ratio between  $z_0$  and  $z_H$  is often defined as (Owen and Thomson 1963):

$$kB^{-1} = \ln \left( \frac{z_0}{z_H} \right) \quad (3)$$

Determining bulk values of  $r_H$  is challenging in urban areas due to the complex geometry and heterogeneous thermal properties of urban materials. Several studies have used observed  $Q_H$ ,  $T_A$ , and  $T_R$  to solve for  $r_H$  using eq. 1. (e.g. Voogt and Grimmond 2000, Kanda et al. 2007, Lee et al. 2016). There are several sources of uncertainty using this approach: i) observed  $Q_H$  obtained from eddy covariance (EC) observations is thought to be typically underestimated by EC given the lack of energy balance closure (Wilson et al. 2001), ii) observations from EC sensors and radiometers are representative of different source areas (Schmid et al. 1991, Schmid 1994), and iii) errors are introduced to  $T_R$  calculations from choice of bulk emissivity values and radiometer bias towards vertically-facing roof and street surfaces (Voogt and Grimmond 2000). Furthermore, the  $r_H$  value is expected to vary in space and time through variability in  $u^*$ ,  $z_0$ , and  $z_H$ . Brutsaert (1982) used a theoretical relation to relate  $kB^{-1}$  (eq. 3) to the roughness Reynold's number ( $Re^*$ ) for rough natural surfaces:

$$kB^{-1} = \alpha Re^{*0.25} - 2.0 \quad (4)$$

where:

$$Re^* = z_0 u^* / \nu \quad (5)$$

where  $\nu$  is kinematic viscosity ( $1.46 \times 10^{-5} \text{ m}^2 \text{ s}^{-1}$ ) and  $\alpha$  is an empirical constant. Kanda *et al.* (2007) experimentally determined an urban value for  $\alpha$  (1.6) using observations over a scale model of bluff body cubes. To our knowledge, only a few studies have used  $LST$  and an aerodynamic resistance approach to derive  $Q_H$  in cities (Kato and Yamaguchi 2005, Weng *et al.* 2014). These studies have produced realistic flux values; however, they are based on relatively few images and lack a measurement network with which to compare results. Overall, there remain challenges with using ARM in urban areas for several reasons: i) uncertainties from satellite-derived measurements from satellite view bias and variable surface bulk emissivity in cities, ii) challenges to resolve spatial variations in air temperature and other meteorological and surface variables needed to calculate resistance in eq. 2, and iii) incomplete theoretical understanding of how  $r_H$ ,  $kB^{-1}$ , and  $z_H$  vary in space and time under different flow and surface forcing conditions.

In this work, we attempt to calculate high-resolution  $Q_H$  based on a large dataset of EO-derived  $LST$  measurements and assess uncertainties with using an aerodynamic resistance approach. First, we analyse spatial  $LST$  patterns and characterize temporal variability of relevant variables, including  $r_H$ ,  $kB^{-1}$ , and  $z_H$ , based on *in situ* observations at an eddy covariance tower. Next, five different variations to calculate  $r_H$  based on observations and statistical relations from the literature are tested and compared. Model uncertainty is evaluated using a Monte Carlo sampling approach of input variables. Central London is used as the test area for developing our methodology because of the presence of a long-term observation network of scintillometers and EC to enable direct comparison of observed and modelled  $Q_H$ .

## 2. Methods

### 2.1 Study domain

The focus of this study is London, UK; the most populous city in Europe (~8.9 million in Greater London Area [GLA, 2017]) and one of three study sites for the H2020 UrbanFluxes project (Chrysoulakis et al. 2015). The spatial domain is a 21.5 km x 21.4 km area encompassing central London (273591.1-295091.1 m E, 5701474-5722874 m N, UTM zone 31, WGS84, Figure 1). All analysis is performed on a 100 m x 100 m pixel basis to conform to Landsat-8 TIRS sensor resolution (Section 2.3).

Physically, the study domain is characterized by high built (building and paved land cover) density in central London near the River Thames and higher vegetation land cover fraction away from the city centre (Figure 1). Overall combined impervious (building and paved) land cover is 45%. Average building heights in the entire domain are around 10 m with areas of high rise buildings in the City of London and Canary Wharf areas with individual building heights up to 310 m.

Vegetation is more abundant outside of the city centre and in several large parks (Figure 1). Overall vegetation land cover ( $\lambda_V$ ) fraction in the domain is 51% (25% grass and shrubs <2.0 m and 26% trees above 2.0 m). Most trees (58%) are deciduous species which follow an annual cycle of leaf-on in northern hemisphere spring (April-May) and senescence and leaf-off in autumn (October-November). Overall, water coverage is 3%, with the city bisected West-East by the River Thames and a system of reservoirs in the Northeast quadrant.

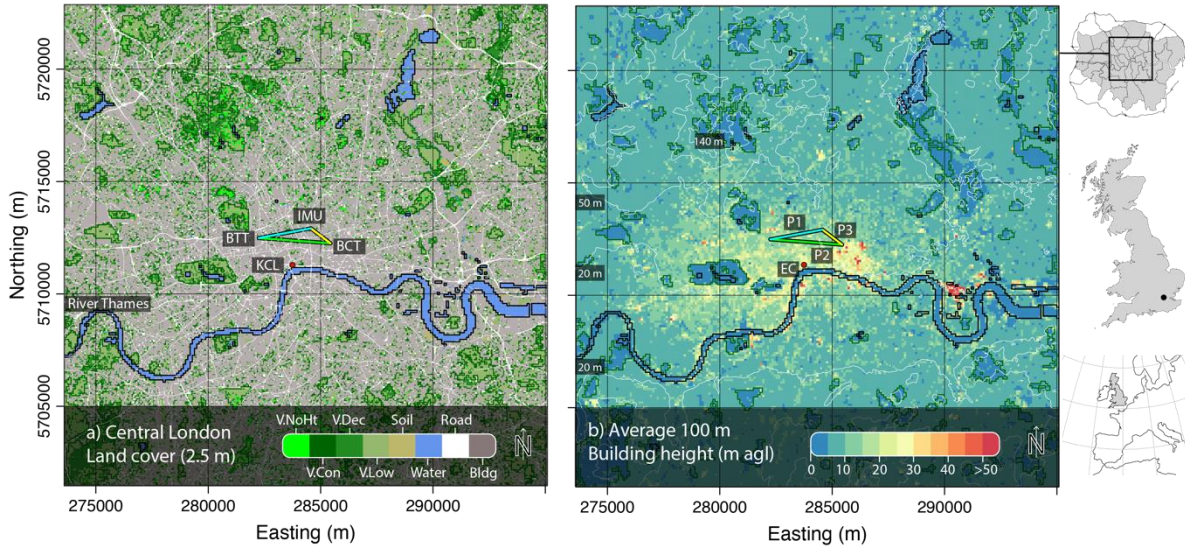


Figure 1. Land cover (2.5 m pixels) and mean building height (100 m pixels) of the study domain with sensible heat flux ( $Q_H$ ) observation network of scintillometers (lines) and eddy covariance (red circle). The significant water bodies (e.g. River Thames) (black) and extensive vegetated areas (e.g. parks) (green) are outlined. Land cover data are derived from multi-spectral Landsat satellite images (2012) and building heights are from a LiDAR digital surface model (Section 2.5). Land cover categories are ‘V.NoHt’ (vegetation without height information), ‘V.Con’ (coniferous vegetation > 2.0 m height), ‘V.Dec’ (deciduous vegetation > 2.0 m height), ‘V.Low’ (vegetation < 2.0 m height), ‘Soil’, ‘Water’, ‘Road’, and ‘Bldg’ (building). Terrain elevation contours measured during the LiDAR scan are overlaid on (b) (30 m intervals, white lines).

## 2.2 London measurement network

A unique feature of this study is the availability of an extensive *in situ* measurement network in central London. Observations from the network are used to develop empirical models of  $r_H$ , provide inputs to calculate  $Q_H$  and turbulent flux source areas, and to evaluate gridded  $Q_H$  calculation results (Figure 1).

Turbulent sensible heat flux and radiation measurements are available from the eddy covariance (EC) tower at King’s College London (KCL). The EC instrumentation consists of a CSAT3 (Campbell Scientific, Logan, UT, USA) sonic anemometer and Li-7500 (Li-Cor, Lincoln, NE, USA) open-path gas analyzer mounted at 60.9 m asl (50.3 m agl). Three-dimensional wind velocities, virtual acoustic air temperature, and water vapour mixing ratios are recorded at 10 Hz and flux processing is performed in 30-minute intervals. Detailed description of the EC site and data processing is found in Kotthaus and Grimmond (2014). Incoming ( $K\downarrow$ ) and outgoing ( $K\uparrow$ ) solar radiation, and incoming ( $L\downarrow$ ) and outgoing ( $L\uparrow$ ) longwave radiation are measured at a height of 60.9 m asl (50.3 m agl) (CNR4 net radiometer, Kipp & Zonen).

A network of three scintillometers provide additional observations of area-averaged  $Q_H$  in the study domain (Table 1, Figure 1). Scintillometers use path-integrated measurements of perturbations to a beam of infrared light to derive  $Q_H$  based on surface layer similarity theory. These instruments provide spatially-integrated measurements from much larger areas than EC. Full description of instrumentation and data processing details for the scintillometer network are found in Crawford et al. (2017).

Satellite overpasses to measure  $LST$  occur between 10:52 and 11:02 UTC and calculated  $Q_H$  at the overpass times is directly compared with simultaneous measurements from the scintillometers and EC system. To ensure measurements are representative of the satellite overpass time, interpolated values between two 30-minute averaged measurement periods (10:30-11:00 and 11:00-11:30 UTC) are used for comparison.

In general, there is uncertainty in the observed-model comparison due to uncertainties with  $Q_H$  observations. Eddy covariance has been shown to have typical hourly uncertainty of 8-10% during daytime periods (Hollinger and Richardson, 2005). This uncertainty comes from a combination of sources including systematic and random sensor errors, vertical flux divergence, and lack of energy balance closure. Scintillometers in complex areas such as central London have typical uncertainty on the order of 20% due to variable meteorological inputs, surface conditions, instrument height, and surface layer similarity theory assumptions (Crawford et al. 2017, Ward 2017).

An additional source of uncertainty in the comparison comes from the use of source area models to weight calculated  $Q_H$  surface means (Heidbach et al. 2017). It is generally understood that source area models do not account for complexities in urban areas resulting from flow channelling in the canopy layer or the influence of heat point sources or plumes (e.g. Crawford and Christen 2014). The individual processes contributing to this uncertainty have not yet been quantified rigorously in an urban setting, but previous sensitivity analysis has shown the scintillometer source area

averaged land cover fractions to be insensitive ( $<\pm 1\%$ ) to changes in source area size of  $-50\%$  to  $+20\%$  in central London (Crawford et al. 2017).

There is also uncertainty because radiometer source areas differ from turbulent EC source areas. This is important because  $r_H$  is calculated using observed  $T_R$  and EC  $Q_H$  at KCL (eq. 1). At KCL, the 90% view factor of the radiometer (50.3 m agl) source area extends to a radius approximately 150 m away from the tower (eq. 3 in Schmid et al. 1991). This extent does not include the River Thames (190 m away from the tower), so the River has only a minimal influence on  $T_R$  measurements. In contrast, the EC measurements have a different and more dynamic source area (i.e. the location changes, not just solar geometry). At times, more vegetation and water are included in the EC source areas, depending on stability and wind direction (over 50% water land cover during neutral conditions with winds from  $175-195^\circ$ , Kotthaus and Grimmond 2014, Kent et al. 2017). This means the River may enhance (depress)  $Q_H$  during night (day) given the relatively warm (cool) water surface, causing a mismatch between observed  $\Delta T (T_R - T_A)$  and  $Q_H$ . This uncertainty is expected to be largest overnight when turbulent conditions are weak and the EC source area dimensions extend further to include the river.

Table 1. Scintillometer and eddy covariance network in central London. The effective beam height is the path-weighted mean height of the beam above the surface, including buildings and trees (Crawford et al., 2017).

	P1	P2	P3	EC
Sensor type	Scintec BLS 900	Kipp & Zonen LAS MkII	Kipp & Zonen LAS MkII	Campbell Scientific, Inc. CSAT 3d sonic anemometer
Effective height (m agl)	111.6	140.7	98.7	50.3
Path (P) length (m)	2358	3197	1097	-

### 2.3 EO-derived LST

Observations from the Enhanced Thematic Mapper Plus (ETM+) and the Thermal Infrared Sensor (TIRS) aboard the Landsat 7 and 8 satellites, respectively, are used to calculate  $LST$  for 26 individual periods from 7 April 2011 – 26 Aug 2016. All scenes have been acquired between 10:52 and 11:02 UTC and processed using the ATCOR algorithm by the German aerospace centre DLR (DLR 2012, Marconcini et al. 2017). As original ETM+ and TIRS thermal bands are 30 m and 100 m resolution, respectively, all  $LST$  products were resampled to 100 m spatial resolution. This procedure also includes image corrections for radiometric, atmospheric, and topographic effects and uses a constant surface emissivity value ( $\epsilon=0.98$ ) (see Section 2.5.4 for discussion of uncertainty related to emissivity). Individual pixels are masked during processing from cloud coverage or instrument malfunction impacts. Images are included in this study if pixel retrieval in the study domain is  $>60\%$ . ETM+ and TIRS sensor field of view is  $15^\circ$  so the viewing angle for individual pixels is within  $\pm 7.5^\circ$  off-nadir.

Landsat-8 images have a known problem of banding due to additional stray light from outside the sensor field of view influencing observations (Montanaro 2014). This additional radiation has been shown to enhance  $LST$  measurements by 1-2 K (in TIRS Band 10). A correction has been developed by Gerace and Montanaro (2017), but this current work was completed before the corrected images became available. However, errors from the stray light problem are considered during uncertainty analysis and results (Section 2.5.4 and Section 3.5-3.6).

### 2.4 Geospatial surface data

Surface morphology parameters (building height [ $z_b$ ],  $z_0$ , and  $z_d$ ) are determined from a high-resolution (1 m) LiDAR-based digital surface model from 2008 (Lindberg and Grimmond, 2011). Morphological variables are calculated for each 100 m pixel and  $z_0$  and  $z_d$  vary according to wind direction (every  $5^\circ$ ) using methods developed by Kanda et al. (2013) as implemented in the UMEP software package (Kent et al., 2017, Lindberg et al. 2018).

Land cover has been classified at 2.5 m based on multi-temporal resolution satellite imagery (SPOT and WorldView) from 2012 and the LiDAR DSM data (Marconcini et al., 2017). Five land cover categories are defined: buildings ( $\lambda_B$ ), paved surfaces ( $\lambda_P$ ), water ( $\lambda_W$ ), bare soil ( $\lambda_S$ ), and vegetation ( $\lambda_V$ ). Vegetation is further categorised as low vegetation ( $<2.0$  m), deciduous vegetation ( $>2.0$  m), coniferous vegetation ( $>2.0$  m), and vegetation with no height information. This 2.5 m resolution land cover data are aggregated to 100 m resolution and land cover fractions range from 0-1 for each category.

### 2.5 Aerodynamic resistance method

#### 2.5.1 Reference height scaling

Gridded geospatial data ( $z_0$ ,  $z_d$ ,  $\lambda_V$ ) and atmospheric variables ( $T_A$ ,  $u_*$ ,  $L$ ) are required at the same resolution as  $LST$  to calculate  $Q_H$  using aerodynamic resistance methods. Atmospheric variables, such as wind speed and  $u_*$ , are height-dependent and respond to surface characteristics of an upwind fetch that typically extend beyond individual 100 m x 100 m grid cells of  $LST$  images. Additionally, corresponding surface variables ( $z_0$ ,  $z_d$ ,  $\lambda_V$ ) for an elevated atmospheric measurement are representative of a larger source area than individual grid cells.

Given the need for appropriate heights for the atmospheric variables and characterisation of the surface, a reference height for the calculations is defined. In this study, a reference height of  $3z_b$  (where  $z_b$  is the mean building height) is

selected. Three is chosen to provide a reasonable likelihood of being above the RSL (Roth 1999, Grimmond and Oke 2000, Grimmond et al. 2004, Kastner-Klein et al. 2004) but recognizing the limitation of the depth the urban surface layer (Rotach 1999).

### 2.5.2 Air temperature and wind fields

Air temperatures from an *in situ* network of meteorological stations are used to generate a  $T_A$  field at 100 m resolution. Currently, observations from a single tower-mounted, shielded, and passively ventilated thermometer (Vaisala, Inc., WXT520) at 50.3 m agl are used across the entire domain. Temperatures are height-adjusted in each pixel to a reference height of  $3z_b$  (Section 2.4) using an environmental lapse rate of  $6.5 \times 10^{-3} \text{ K m}^{-1}$ .

During model development,  $T_A$  fields were interpolated from distributed networks of air temperature observations (e.g. NetAtmo, WeatherUnderground). However, these datasets introduce additional uncertainty in the current application because of i) uneven spatial distribution of sensors, ii) lack of reliable location metadata for many stations, and iii) lack of evaluation with co-located observations (Gabey et al. 2018). Other  $T_A$  measurements in central London are available from the London Urban Meteorological Archive (LUMA), but no consistent set of stations are available for the entire 2011-2016 study period. So, to ensure consistency across the study period, we opted to use a single, high-quality measurement (KCL) as  $T_A$  reference at the expense of higher spatial resolution. This procedure introduces some uncertainty because we are unable to resolve phenomena such as urban heat or cool islands (Theeuwes et al. 2015).

This simplification is justified because spatial variability of  $T_A$  across the study domain at the reference height ( $3z_b$ ) is expected to be much less than that of  $LST$ . This means  $T_A$  is essentially spatially constant relative to  $LST$  and that spatial variations in  $LST$  dominate the  $(T_A - LST)$  term in Eq. 1. To further assess uncertainty with  $T_A$ , other long-term monitoring stations (e.g. airports) located outside the study domain are also used in Section 2.5.4.

Gridded  $u$  and  $L$  values are required as inputs to Eqs. 1-2 to determine  $r_H$ . Results from a sensitivity analysis (Section 3.5) show that  $Q_H$  is relatively insensitive to  $L$ , so  $L$  is taken from KCL sonic observations and assumed spatially constant across the domain. The gridded  $u$  field is determined by the following procedure: Observed wind speed ( $U$ ), friction velocity ( $u_*$ ), and Obukhov length ( $L_{KCL}$ ) are measured by the KCL sonic anemometer ( $z_{m, KCL} = 50.3 \text{ m agl}$ , approximately  $3z_b$  (2012-2016)). The corresponding roughness parameters  $Z_{0, KCL}$  and  $Z_{d, KCL}$  (Section 2.4) are used in the stability-adjusted ( $\psi$ ) logarithmic wind profile (using the sign convention of positive  $\psi$  as in Stull, 2012):

$$U = \frac{u_*}{k} \left[ \ln \left( \frac{z_{m, KCL} - Z_{d, KCL}}{Z_{0, KCL}} \right) + \psi \left( \frac{z_{m, KCL} - Z_{d, KCL}}{L_{KCL}} \right) \right] \quad (6)$$

The measured wind speed is then extrapolated to an arbitrary reference height of 200 m asl ( $z_{ref}$ , selected to be frequently well above the RSL and reasonably assume a spatially homogeneous reference wind speed,  $U_{ref}$ ) assuming vertically constant  $u_*$ :

$$U_{ref} = U \frac{\left[ \ln \left( \frac{z_{ref} - Z_{d, KCL}}{Z_{0, KCL}} \right) + \psi \left( \frac{z_{ref} - Z_{d, KCL}}{L_{KCL}} \right) \right]}{\left[ \ln \left( \frac{z_{m, KCL} - Z_{d, KCL}}{Z_{0, KCL}} \right) + \psi \left( \frac{z_{m, KCL} - Z_{d, KCL}}{L_{KCL}} \right) \right]} \quad (7)$$

The standardised reference wind speed  $U_{ref}$  is then applied across the domain and scaled to the  $3z_b$  reference height for each grid cell to calculate pixel-specific  $U'$ . This uses the roughness parameters estimated for a particular location and assumes spatially consistent stability (i.e.  $L$ ) across the domain:

$$U' = U_{ref} \frac{\left[ \ln \left( \frac{3z_b - Z_d}{Z_0} \right) + \psi \left( \frac{3z_b - Z_d}{L_{KCL}} \right) \right]}{\left[ \ln \left( \frac{z_{ref} - Z_{d, KCL}}{Z_{0, KCL}} \right) + \psi \left( \frac{z_{ref} - Z_{d, KCL}}{L_{KCL}} \right) \right]} \quad (8)$$

### 2.5.3 Aerodynamic resistance parameter

The aerodynamic resistance term ( $r_H$ ) is calculated (eq. 2) using MOST. It varies both temporally and spatially. In this work, five approaches to calculate  $r_H$  are compared (Table 2, also see Appendix Table A.2). Three approaches use empirical relations derived for the central London KCL flux site and two are generalized statistical relations for urban areas based on scale model and urban flux measurements. Additional methods of Jurges (1924), Rowley (1930), and Harman (2004) were also tested preliminarily, however these methods all require a parameterization of high-resolution canopy layer wind speed which is beyond the scope of this work to generate with high confidence.

Table 2. Five methods to determine aerodynamic resistance in central London on a pixel basis.

Method	Method source	Description
R1	This paper	Constant $r_H$ , based on observations at KCL (eq. 1)
R2	This paper	Empirical relation between $z_H$ and solar elevation ( $\delta$ )
R3	This paper	Empirical relation between $z_H$ and $u$
R4	Kanda et al. (2007)	Empirical relation between $kB^{-1}$ and $Re^*$ at urban scale model (eq. 4)
R5	Kawai et al. (2009)	Empirical relation between $\alpha$ and land cover in real cities (eq. 8)



### 2.5.3.1 R1-R3 from central London observations

Three methods rely on findings from KCL tower (central London) observations. These are described from simplest to most complex. The empirical relations are derived using daytime observations ( $K_d > 10 \text{ W m}^{-2}$ ), non-stable ( $EC Q_H > 20 \text{ W m}^{-2}$ ), dry surface conditions (time since precipitation  $> 12$  hours) to correspond to general conditions during satellite image acquisition. The observation results for these methods are in Section 3.2-3.3.

The R1 method uses  $r_H$  calculated from Eq. 1 and observed  $Q_H$ ,  $T_A$ , and substituting radiometric surface temperature ( $T_R$ ) for  $T_0$ . The  $T_R$  is calculated using the Stefan-Boltzmann equation with observed longwave radiation upward and downward components and an assumed bulk emissivity of 0.95 for an urban setting (Arnfield 1982, Kawai et al. 2009, Kotthaus et al. 2014). In general, there is uncertainty in using  $T_R$  as a proxy for  $T_0$  arising from i) the choice of emissivity value, ii) lack of atmospheric correction, iii) source area differences between the radiometer and eddy covariance instruments, and iv) radiometer view bias.

The mean  $r_H$  value of  $40 \text{ s m}^{-1}$  for all periods in 2015 for 10:30-11:30 UTC is used (approximate time of satellite overpass) for all images. This is a simple approximation that assumes spatially constant  $r_H$  and does not account for variations in atmospheric conditions ( $u^*$ ,  $L$ ) nor surface land cover and roughness spatial variations.

The R2 method calculates  $z_H$  based on an empirical relation between solar elevation ( $\delta$ , ° above horizon) and  $z_H$  observed at the central London eddy covariance tower (Section 3.3). The  $z_H$  is found by first solving for  $r_H$  using Eq. 1 and then solving for  $z_H$  in Eq. 2, using available tower-based (50.3 m agl) meteorological observations ( $u^*$ ,  $L$ ,  $T_A$ ), surface parameters ( $z_0$  and  $z_d$ ) calculated in  $10^\circ$  wind direction increments (Kanda et al. 2013, Crawford et al. 2017), and substituting radiometric surface temperature ( $T_R$ ) for  $T_0$ . Solar elevation is calculated for the day of year and time of each overpass (Meeus 1991, NOAA 2017).

The R3 method uses  $z_H$  determined from a linear relation between mean wind speed ( $U$ ) and  $z_H$  observed at the KCL tower (Section 3.3). The  $z_H$  is calculated as described for R2 and wind speed varies spatially (Section 2.5.2) depending on surface roughness (Section 2.4) and reference height (Section 2.1).

For R2-R3, the  $r_H$  value is calculated for each pixel using Eq. 2 once  $z_H$  has been determined. The stability function for heat ( $\psi_h$ ) used for unstable conditions in Eq. 2 is (van Ulden and Holtslag, 1985):

$$\psi_h = 2 \ln \left( \frac{1+x^2}{2} \right) \quad (9)$$

where (Högström, 1988):

$$x = 0.95 \left[ 1 - 15.2 \left( \frac{z'}{L} \right) \right]^{0.5} \quad (10)$$

where  $z' = z_m - z_d$  and  $z_m$  is the measurement height.

The  $u^*$  and  $L$  values in Eq. 2 are assumed spatially constant and set to observed values at KCL. This simplification is necessary to make the problem more tractable, though the assumption of spatial homogeneity is likely not physically realistic across this study domain and artificially decouples  $Q_H$  from spatial variations in  $L$  and  $u^*$  (e.g. Su 2002). Our approach instead is to quantify uncertainty resulting from this simplification (Section 3.5).

During model development, an iterative approach using linked, spatially dynamic  $u^*$ ,  $L$ , and  $Q_H$  values was explored. However,  $u^*$  and  $L$  values were not well-constrained and this iterative procedure resulted in greater model uncertainty and poorer comparison to observations than the simpler approach. Overall,  $Q_H$  results are found to be relatively insensitive to variations in  $L$  and more sensitive to  $u^*$  (Section 3.5).

### 2.5.3.2 Literature based methods (R4-R5)

Empirical relations from observations at KCL (R1-R3) are compared to established methods described in the urban climate literature. The R4 method uses Kanda et al.'s (2007) 'urban'  $\alpha$  parameter (1.29) for Eq. 4 based on observations over the COSMO urban scale model in Japan (Figure 2). The difference from the  $\alpha$  value used for rough-bluff natural surfaces (2.46) is attributed to differences in mean flow regimes and different surface roughness of the urban model compared to bluff natural surfaces (Kanda et al. 2007). It is important to note that i) the COSMO model has neither vegetation nor anthropogenic heat ( $Q_F$ ) sources (both will influence  $kB^{-1}$  and observed  $Q_H$  values) and ii) the original R4 model formulation is designed for non-vegetated urban surfaces. Therefore, R4 results are expected to differ from R1-R3 (empirical models tuned for this specific site). R4 inclusion is intended as a relevant benchmark; therefore, 'Observed - R4' differences provide potential constraints on the influence of  $Q_F$  and vegetation.

During implementation,  $Re^*$  is determined based on  $u^*$  and  $z_0$  for each pixel. For pixels in the London domain with water or vegetation land cover fractions  $> 0.8$ , the  $\alpha$  value is set to 2.46 of 'natural' rough-bluff surfaces.



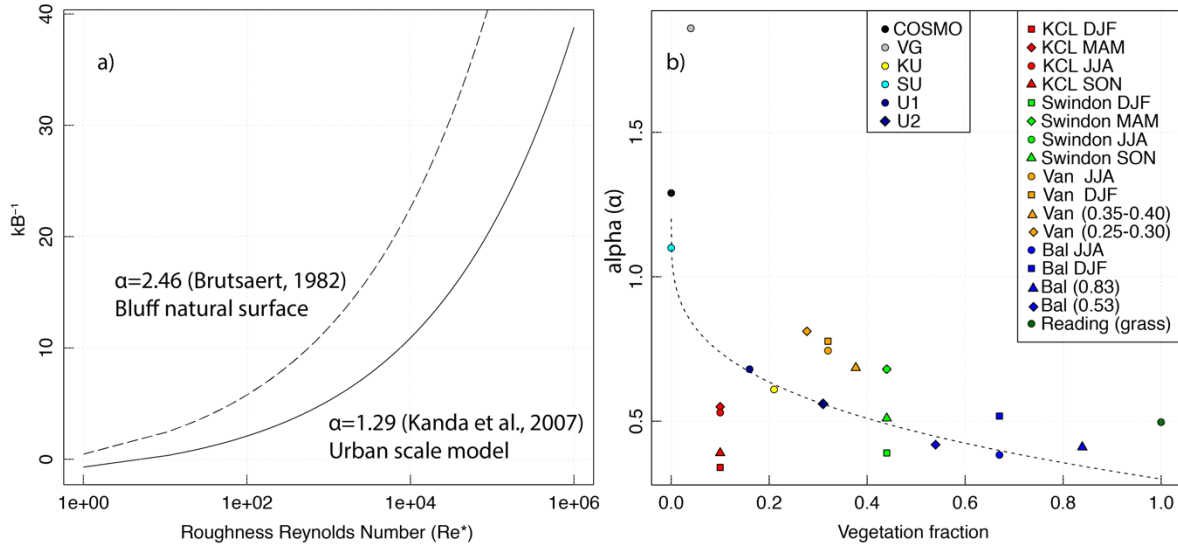


Figure 2. Statistical relations from a) physical urban scale model (COSMO, Kanda et al., 2007) and b) local-scale urban observation sites (Kawai et al. 2009) used to calculate the aerodynamic resistance term ( $r_H$ ). For b) ‘COSMO’ scale model (Kanda et al. 2007), Vancouver-light industrial (‘VG’, Voogt and Grimmond 2000), Tokyo Kugahara (‘KU’, Moriwaki and Kanda 2006), Tokyo business district (‘SU’, Sugawara 2001), Basel-Sperrstrasse (‘U1’), and Basel-Spalenring (‘U2’) (Christen and Vogt, 2004) are those used by Kawai et al. (2009). Results from central London (‘KCL’, Kotthaus and Grimmond 2014), Swindon (Ward et al., 2013), Vancouver-Sunset (‘Van-ST’, Crawford and Christen, 2015), Baltimore (‘Bal’, Crawford et al. 2011), and University of Reading reference grass site (‘Reading’) are added in this work. These sites are shown for contrasting meteorological seasons (DJF, MAM, JJA, SON), and average summertime vegetation land cover fractions.

The R5 method allows  $r_H$  to vary spatially with changes in vegetation land cover ( $\lambda_v$ ) as proposed by Kawai *et al.* (2009) (Figure 2b). The original findings use data from five local-scale urban EC measurement campaigns with a vegetated land cover fraction range of 0-0.3. The  $\alpha$  value in Eq. 4 is found to decrease as vegetation land cover fraction increases:

$$\alpha = a_u - (a_u - a_L) * \lambda_v^b \quad (11)$$

where  $a_u$  (1.2) and  $b$  (0.29) are empirical coefficients and  $a_L$  is set to 0.3 based on  $\alpha$  found experimentally in completely vegetated ecosystems. The R5 method permits the urban fraction within grids of numerical weather prediction to vary. The application here is an extension beyond the original purpose to determine if i) the observed relation holds for vegetation land cover fraction values greater than 0.3 and ii) whether this can be applied at higher resolutions and evaluated against measurements in real cities.

Here, the urban vegetation land cover range is expanded to 0-0.8 by adding four additional sites: Vancouver-Sunset, Canada (Crawford and Christen, 2015); London-KCL, UK (Kotthaus and Grimmond, 2014); Swindon, UK (Ward et al. 2013); and Baltimore, USA (Crawford et al. 2011).

For each site, the  $\alpha$  value is found by rearranging Eqs. 1-2 to solve for  $kB^{-1}$  and then plotting  $kB^{-1}$  against observed  $Re^*$ . At each site, meteorological variables ( $u^*$ ,  $L$ ,  $T_A$ ) are observed,  $T_R$  is calculated and used as a substitute for  $T_0$  (Section 2.5.1), and surface parameters ( $z_0$  and  $z_d$ ) are determined for  $20^\circ$  wind sectors using anemometric methods in neutrally stable conditions (Grimmond and Oke, 1999). Additionally, data are stratified by season (DJF, MAM, JJA, SON) to assess variability from changing canopy and weather conditions and by wind sector (for Baltimore and Vancouver-Sunset) to distinguish between source areas with contrasting vegetation land cover fractions.

The overall trend described by Kawai *et al.* (2009) is supported by the additional observations (Figure 2b), though there is scatter about the trend line (mean absolute deviation =  $\pm 0.18$ ). A physical explanation for this trend is that momentum transfer is enhanced relative to heat transfer due to increased pressure drag from dominance of bluff bodies in less vegetated urban environments. This increases  $\alpha$ , thus also  $kB^{-1}$  and  $r_H$  values. As vegetated land cover fraction increases, the canopy becomes more porous, pressure drag is reduced, and heat transfer efficiency is enhanced by the intricate vegetation canopy 3D surface. Values for  $\alpha$  determined experimentally for rough, porous vegetated canopies such as pine forest and savannah scrub are approximately 0.3 (Kawai et al. 2009 and references therein).

London-KCL appears as an outlier in this figure with lower  $\alpha$  values than expected (0.45) based on vegetation fraction (0.09). A possible explanation for this could be that closely packed buildings in central London place this site more towards the skimming flow regime where form drag from bluff buildings is diminished. Another explanation is the influence of the River Thames on EC measurements and differing source areas between  $T_R$  (radiometer) and  $Q_H$  (EC) used to derive  $r_H$ ,  $kB^{-1}$ , and  $\alpha$  (Section 2.2).

To implement R5, pixel-specific  $\alpha$  is derived using  $\lambda_v$  and Eq. 11. Then  $kB^{-1}$  is calculated using Eq. 4. For water or vegetation land cover fractions  $>0.8$ , the  $\alpha$  value is set to 2.46 of ‘natural’ rough surfaces.

## 2.5.4 Quantifying uncertainty

To quantify model uncertainty, a modified Monte Carlo approach is used to sample the parameter space of input meteorological and surface variables. The model is run with 20 unique sets of inputs within a prescribed range from the original observed or calculated value (Table 3). Values within this range are generated using quasi-random numbers. For each run, each variable is sampled at a unique location in the allowed range. The quasi-random numbers are generated assuming a uniform distribution and this conceptually represents a systematic bias which affects all individual grid cells (i.e. individual grid cell uncertainties are not independent).

The range for each variable was determined using a variety of approaches. For  $LST$ , we consider three independent sources of error: surface emissivity ( $\varepsilon$ ), sensor bias, and satellite view bias. For  $\varepsilon$ , the uncertainty interval is determined by calculating  $T_R$  from KCL radiometer measurements with bulk emissivity set to 0.9 and 1.0 as physically reasonable upper and lower constraints. The resulting sensitivity of  $T_R$  to this  $\varepsilon$  range is on average 1.2 K ( $\pm 0.6$  K) for all hours at 11:00 UTC from 2012-2016. The KCL  $T_R$  during each overpass (with  $\varepsilon=0.98$ ) is compared directly to extracted  $LST$  values (weighted mean of pixels surrounding the KCL tower to approximate the radiometer source area) and  $LST$  is on average 0.8 K higher than  $T_R$ . This is comparable to the  $LST$  bias (0.67 K) found between application of constant and variable emissivity values to multispectral satellite images in an urban area (Mitraka et al. 2012).

The sensor bias is the reported average absolute error estimate due to the stray light problem (+1 to +2 K) (Gerace and Montanaro 2017). For satellite view bias,  $LST$  is likely underestimated relative to the complete surface temperature by 0 to -2.5 K at 11:00 UTC (which acts to offset the stray light sensor bias), if we assume a nadir viewing angle. This is based on thermal camera measurements in an urban neighbourhood (Adderley and Christen 2015) and a physical scale urban model (Roberts 2010) that found maximum differences between nadir and complete surface temperatures at solar noon (2.2-2.5 K) and near-zero difference at 10:30 local time (Adderley and Christen 2015). This error is expected to vary spatially and generally increase with building height and canyon aspect ratio as the proportion of walls to the total 3D urban surface area changes (Krayenhoff and Voogt, 2016). Additional errors result from the actual off-nadir view angle ( $\pm 7.5^\circ$ ), satellite azimuth angle, and thermal anisotropy of the surface. There are likely further sources of error in  $LST$  observations, such as neglecting urban aerosols during atmospheric correction, that are not quantified.

For  $T_A$ , the uncertainty range is the mean difference between KCL and Heathrow airport on the outskirts of central London for 10:30-11:30 UTC; this is taken to be the likely maximum air temperature difference within the spatial domain during the overpass times. For wind speed, the average difference between predicted wind speeds using the methods described in Section 2.5 and actual measurements from two stations in the London network (BCT, IMU) are used. For  $z_0$  and  $z_d$ , the range is based on analysis of several morphometric and anemometer-based methods to determine their values in central London (Kent et al. 2017). For  $\alpha$  in Eq. 9, the value is the mean absolute difference between the fit model curve and observation points in Figure 2b. The ranges of  $u^*$  and  $L$  are set to the standard deviation of measurements at the KCL flux tower at 10:30-11:30 UTC during an entire year (2016), essentially substituting observed temporal variability for spatial variability. The solar elevation uncertainty is based on variations in overpass time (10:52-11:02 UTC) and from simplified orbital calculations.

Vegetation fraction uncertainty is set to the 100 m pixel difference ( $\pm 0.1$ , includes 95% of 100 m pixels) between land cover determined for this project and that from the UK Ordnance Survey MasterMap (Ordnance Survey 2017) in a subset of the modelling domain. For elevation, uncertainty is estimated from a comparison between LiDAR-derived elevation datasets from 2008 and 2015 for a spatial subset where there is overlap between the datasets. LiDAR instrumentation errors are  $< \pm 1$  m and uncertainty primarily is from changes due to building construction. Between 2008 and 2015, 99% of 100 m pixels are within  $\pm 4$  m elevation difference (though roughness element heights are on average 0.5 m taller in 2015 compared to 2008).

It should be noted that there are likely other sources of uncertainty in individual pixels resulting from process that have not been quantified. For example, complexities related to sub-pixel heterogeneity are not considered explicitly and advection from cool grid cells (e.g. parks, rivers) to warmer adjacent pixels could enhance micro-scale sensible heat fluxes by increasing micro-scale vertical temperature gradients.

Table 3. Input variables to calculate  $Q_H$  and their estimated uncertainty range used for the Monte Carlo and sensitivity analyses.

Variable	Description	Estimated uncertainty range ( $\pm$ )	How uncertainty range is determined
$LST$	EO-observed from Landsat-8 satellite based on thermal infrared radiation	2 K	Combined uncertainty from: i) surface emissivity, ii) sensor bias, iii) view-angle bias.
$T_A$	Air temperature spatially interpolated from observation network	0.5 K	Mean difference between central London and Heathrow airport at 11:00 UTC.
$U$	Reference mean wind speed observed at KCL tower, spatially calculated using log law	3 m s <sup>-1</sup>	Mean difference between predicted and observed values at 2 stations in central London.
$u^*$	Friction velocity observed at KCL tower, assumed spatially constant	0.15 m s <sup>-1</sup>	IQR of measured values for a full year at KCL at 11:00 UTC
$z_d$	Displacement height calculated from DSM using Kanda et al. (2013) method, varies by wind direction	10 m	Based on analysis of morphometric and anemometer-based approaches in central London (Kent et al. 2017)

$z_0$	Roughness length calculated from DSM using Kanda et al. (2013) method, varies by wind direction	2 m	Based on analysis of morphometric and anemometer-based approaches in central London (Kent et al. 2017)
$z$	Reference height determined LiDAR observed DSM from 2008	4 m	Comparison between 2008 and 2015 LiDAR elevation datasets.
$L$	Obukhov Length observed at KCL tower, assumed spatially constant	300 m	IQR of measured values for a full year at KCL at 11:00 UTC
$\lambda_v$	Vegetation land cover fraction from remote sensing	0.1	Estimate based on comparison between two land cover datasets.
$\alpha$	Alpha value in Eqs. 4 and 8 to calculate $kB^{-1}$	0.18	Mean absolute deviation of observed values about the model curve in Figure 2b.
$\delta$	Solar elevation angle	3°	Based on variations in overpass time and reference time of 11:00 UTC and orbital simplifications in solar model.

Thus 20 ensemble members are generated for each pixel and each method, for each  $LST$  image. The uncertainty range for each method is estimated from the median absolute difference between the 10% and 90%<sup>ile</sup>  $Q_H$  values calculated for each pixel. The median value of this matrix is then interpreted as the average uncertainty range for each method across the spatial domain (both absolute  $W\ m^{-2}$  and relative to median  $Q_H$  for all images).

### 3 Results

#### 3.1 EO-derived $LST$

Median  $LST$  in the study domain derived from EO measurements follows an expected annual course with highest values in summer and lowest in winter (Figure 3a). The median  $LST$  values for each scene are generally higher than the  $T_R$  range (5<sup>th</sup>-95<sup>th</sup> %<sup>ile</sup>) observed at KCL from 2012-2016 because the EO scenes are from cloud-free periods and do not include overcast and rainy times.  $LST$  values are also all warmer than air temperatures, implying unstable atmospheric conditions and positive  $Q_H$ .

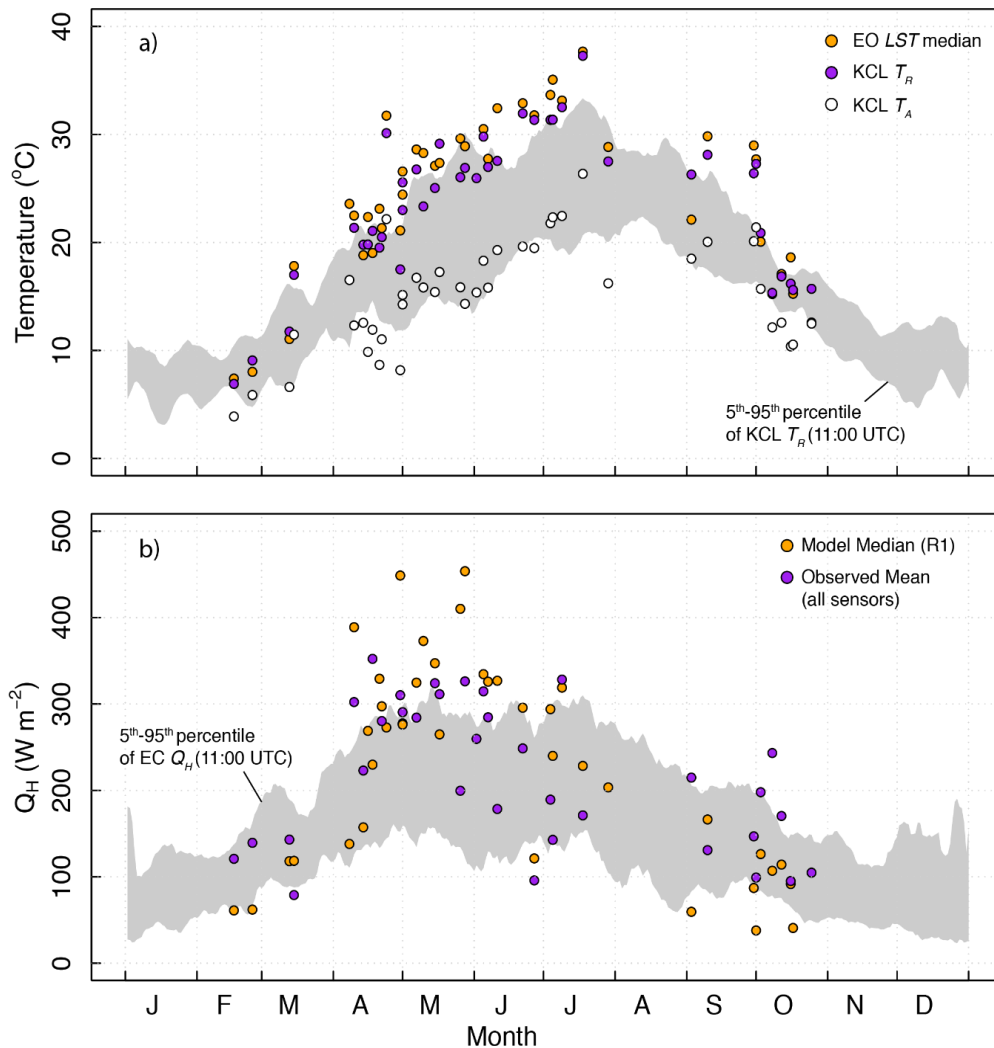


Figure 3. Annual course of a) EO-derived  $LST$ , tower-observed radiometric temperature ( $T_R$ ) and air temperature ( $T_A$ ) and b) tower-observed and modelled  $Q_H$  (R1). Observed  $T_R$  and  $T_A$  in a) and observed and modelled  $Q_H$  in b) are at the same time as each  $LST$  overpass. The 5<sup>th</sup>-95<sup>th</sup> %<sup>ile</sup> ranges of observed  $T_R$  values (grey shading) is shown for all days from 2012-2016 at 11:00 UTC with a 10-day smoothing filter. For observed  $T_R$  (a), the range of values includes emissivity variations from 0.90-1.0. For  $Q_H$  (b), observed values are from the three scintillimeters and EC system (Figure 1).

Spatially, coolest temperatures are generally found in more vegetated areas and for the surface of water bodies.  $LST$  tends to increase with higher building land cover fraction ( $\lambda_B$ ) (Figure 4a). The greatest rate of increase in  $LST$  is when  $\lambda_B$  ranges from 0-0.5, above which  $LST$  remains constant (5-7% above spatial mean). In winter,  $LST$  is relatively high (10% above mean) in areas with high building fractions ( $>0.8$ ), possibly due to anthropogenic heating, however there are few data points from winter ( $n=2$ ). In this analysis, pixel  $LST$  is normalized relative to the spatial mean ( $\langle LST \rangle$ ) for each scene ( $LST/\langle LST \rangle$ ) and only pixels with water land cover fractions  $<0.5$  are used. The normalized  $LST$  pixels are binned in land cover fraction increments of 0.1. The relation between vegetation land cover and  $LST$  is the inverse (not shown), i.e.  $LST$  decreases as vegetation fraction increases.

Though  $LST$  increases as building land cover increases, measured  $LST$  tends to be cooler in areas with tall buildings (Figure 4b). Pixels with tallest buildings ( $>120$  m on average) are in general 10% cooler than the spatial average. In this analysis, only pixels with vegetation and water land cover fractions  $<0.5$  are selected and normalized  $LST$  pixels are binned in 10 m building height increments.

This cooling trend is evident for various classes of building land cover fractions. For example,  $LST$  is relatively low when  $\lambda_B < 0.2$ , yet  $LST$  still decreases as mean building height increases. Another notable trend is that when buildings are shorter ( $<80$  m),  $LST$  increases with increasing  $\lambda_B$ , consistent with Figure 4a. However, when buildings are taller ( $>110$  m), this pattern reverses and  $LST$  tends to decrease with  $\lambda_B$ .

When buildings are shorter, the positive correlation between  $LST$  and  $\lambda_B$  can intuitively be explained due to the greater fraction of buildings and their rooftops at the expense of cooler grass and tree vegetation, with a satellite view bias towards warm rooftops (at the 11:00 UTC overpass time). In the tall building regime, the apparent cooling as building density increases could result from several explanations: i) increased surface shading and reduced sky view factor from tall buildings, ii) increased proportion of shaded walls that are included in off-nadir satellite view, iii) reduced emissivity of building materials such as glass and steel present in higher proportions in areas with tallest buildings.

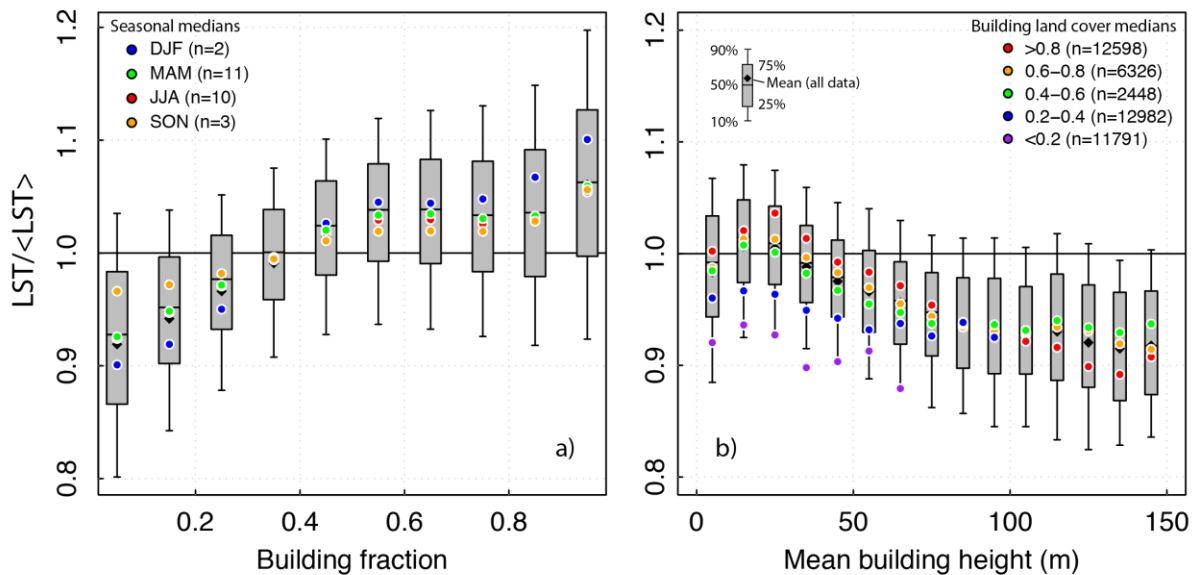


Figure 4. Normalized  $LST$  for all available Landsat images binned by a) building land cover fraction with medians for meteorological season (coloured dots) and b) mean building height with medians for building land cover fraction (coloured dots). In a) only pixels with water  $<0.5$  are used and in b) only pixels with vegetation and water  $<0.5$  are selected.

Four images from contrasting periods (summer day, summer night, winter day, and winter night) are chosen as representative examples for comparison. Based on these images, daytime  $LST$  has higher spatial variance (s.d.=2.67 K on 8 June 2013 11:00 UTC) whereas nighttime  $LST$  is more uniform (s.d.=0.12 on 8 June 2011 21:43 UTC) (Figure 5). During winter (not shown), daytime variability is less than in summer (s.d.=1.08 K on 16 Feb 2016), though nighttime variability between seasons is similar (s.d.=0.13 K on 25 Dec 2011 21:44 UTC).

$LST$  patterns also exhibit spatial structure correlated with vegetation and building land cover fraction. Variograms for the four contrasting periods show spatial autocorrelation across a range of spatial scales (Figure 6). Low variance is associated with homogenous surfaces and high variance with heterogeneity.

Summer daytime has highest variance at all scales, while summer and winter nights have the lowest. For all images however, variance increases up to 500 m where there is an inflection and the variance curve becomes relatively flat. This implies the spatial surface temperature structures are more homogenous (i.e. less variance) at length scales less than 500 m. In contrast, a field of uniformly distributed random numbers has a flat variance curve across the entire range of length scales. Vegetation and building land cover fractions also exhibit an inflection at around 500 m, suggesting observed  $LST$  patterns are linked to the built and vegetated structure in the area.



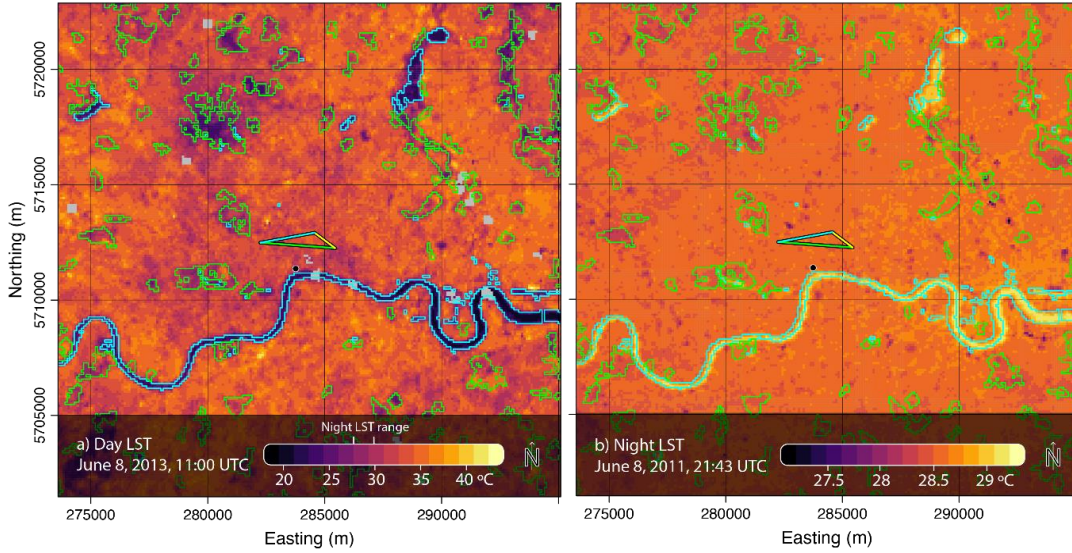


Figure 5. Summer EO-derived  $LST$  for representative a) day and b) night periods. Daytime images are from Landsat satellite and nighttime images are from the ASTER satellite. Parks (green), water bodies (blue), and the central London  $Q_H$  observation network are also shown (coloured lines and black circle).

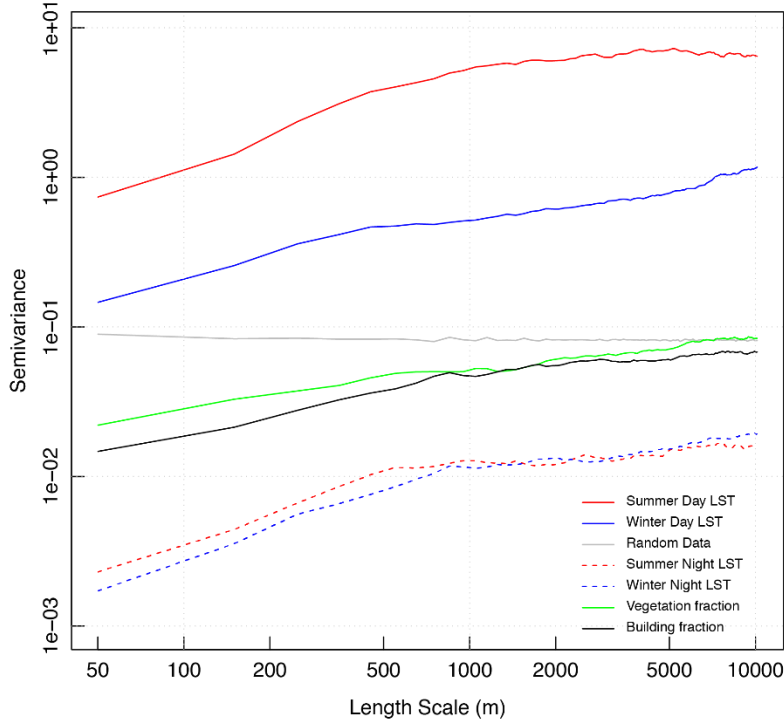


Figure 6. Variogram of  $LST$  from four contrasting time periods. A field of uniformly distributed random data is shown for comparison. Both axes are logarithmic.

### 3.2 Temporal variability of $r_H$

To assess temporal variability of  $r_H$  and variables used to calculate  $r_H$ , we use continuous observations from the KCL EC tower for a full year (2015). Observed  $Q_H$  and  $\Delta T$  ( $T_R - T_A$ ) follow expected diurnal and seasonal courses, with highest values during spring-summer (MAM-JJA) and midday periods and lowest values during winter (DJF) and overnight (Figure 7a-b).  $Q_H$  tends to be highest during spring (MAM,  $>200 \text{ W m}^{-2}$  from 1100-1500 h), though  $\Delta T$  is highest during summer ( $>8^\circ\text{C}$  from 1000-1300 h). This is consistent with prior observations at this site that found highest  $Q_H$  in springtime when there is relatively plentiful energy input from  $Q^*$ , but vegetation evapotranspiration is not yet at full strength to partition energy towards  $Q_E$  (Kotthaus and Grimmond 2014). For Figure 7a-f, daytime ( $K\downarrow > 10 \text{ W m}^{-2}$ ) data from unstable periods ( $Q_H > 20 \text{ W m}^{-2}$ ) with a dry surface (time since rain  $> 12 \text{ h}$ ) in 2015 are used.

The thermal resistance ( $r_H$ ) calculated (Eq. 1) at KCL for all hours in 2015 using measurements of  $Q_H$ ,  $T_A$ , and  $T_R$  ( $\varepsilon=0.95$ ) has a clear diurnal pattern (Figure 7c). The highest median  $r_H$  values occur during midday ( $36.9 \text{ s m}^{-1}$  from 1000-1400 h) and lowest overnight (median,  $10.0 \text{ s m}^{-1}$  from 2100-0500 h). There is considerable variability (mean of the s.d. for all binned hours is  $52.4 \text{ s m}^{-1}$ ). Seasonally, median values are highest in summer (peak at  $55.6 \text{ s m}^{-1}$  at 1100 h) and lowest during winter (peak  $23.1 \text{ s m}^{-1}$  at 1200 h). Seasonal differences in  $r_H$  are likely due to a combination of seasonal

differences in vegetation canopy state and anthropogenic heat flux ( $Q_F$ ). Constant  $r_H$  at 10:30-11:30 ( $40 \text{ s m}^{-1}$ ) is used for the R1 method to model  $r_H$  (Section 2.5).

It is somewhat counterintuitive that  $r_H$  is largest during the day, when  $Q_H$  and  $\Delta T$  are highest. This differs from the diurnal pattern when  $r_H$  is modelled directly using MOST (Eq. 2) with  $kB^{-1}$  from eq. 4 ( $\alpha$  set to 0.45, Figure 2b). The modelled diurnal cycle shows  $r_H$  is lowest during midday when  $Q_H$  and  $u^*$  are highest, and highest overnight (also see Appendix 1 Figure A1, Table A2). During daytime, modelled and ‘observed’  $r_H$  values are comparable, however, differences occur overnight (up to  $30 \text{ s m}^{-1}$ ). A reason for this discrepancy could be due to source area differences between the radiometer and turbulent EC source areas at KCL, particularly overnight (Section 2.2). Though this undoubtedly introduces uncertainty to estimates of  $r_H$  from measurements at KCL, the same diurnal pattern of  $r_H$  as Figure 7 is also found when using both i)  $Q_H$  from the P2 scintillometer to solve for  $r_H$  (along with  $T_A$  from IMU and  $T_R$  from KCL), which is affected much less by the river and ii) EC  $Q_H$  filtered by wind direction ( $300^\circ - 45^\circ$ ) to avoid the river’s influence. Additionally, there is overall greater uncertainty during nighttime both because EC measurements have greater uncertainty due to less vigorous vertical mixing and MOST assumptions begin to break down (e.g. Best and Grimmond 2015).

Magnitudes of  $r_H$  observed at the KCL tower are slightly higher than observed from the scintillometer source areas. The scintillometer  $r_H$  is estimated using observed  $Q_H$ ,  $T_A$ , and scintillometer source area-weighted  $LST$  in Eq. 2. Mean values for satellite overpass times range from  $32.6 - 39.2 \text{ s m}^{-1}$  (-18.5% to -2.0% difference from KCL). Differences are likely due to land cover and urban form differences between scintillometer and EC source areas, differences between  $LST$  and  $T_R$ , and from satellite overpass bias towards clear-sky conditions.

The same diurnal and seasonal cycles as  $r_H$  are evident in hourly median  $kB^{-1}$  values for 2015 at KCL (Figure 7d). The  $kB^{-1}$  term is calculated using Eq. 2 with  $r_H$  determined from observations (Eq. 1), measurements from KCL ( $u^*$ ,  $L$ ), and  $z_0$  and  $z_d$  calculated for  $10^\circ$  wind sectors using the Kanda et al. (2013) method (Section 2). Physically, this can be interpreted as enhanced momentum transfer relative to heat transfer during midday. Momentum transfer is enhanced by form drag in urban areas, while heat transfer is influenced more by molecular diffusion near the surface and thus tends to encounter greater aerodynamic resistance (Kanda et al. 2007). This interpretation is also supported by observation of higher  $u^*$  and lower  $z_H$  values during afternoon. We have higher confidence in these diurnal patterns of  $kB^{-1}$  and  $z_H$  than in  $r_H$  because they generally agree with modelled diurnal cycles using Eqs. 3-4 (though the modelled range of values is much smaller than measurement-derived values).

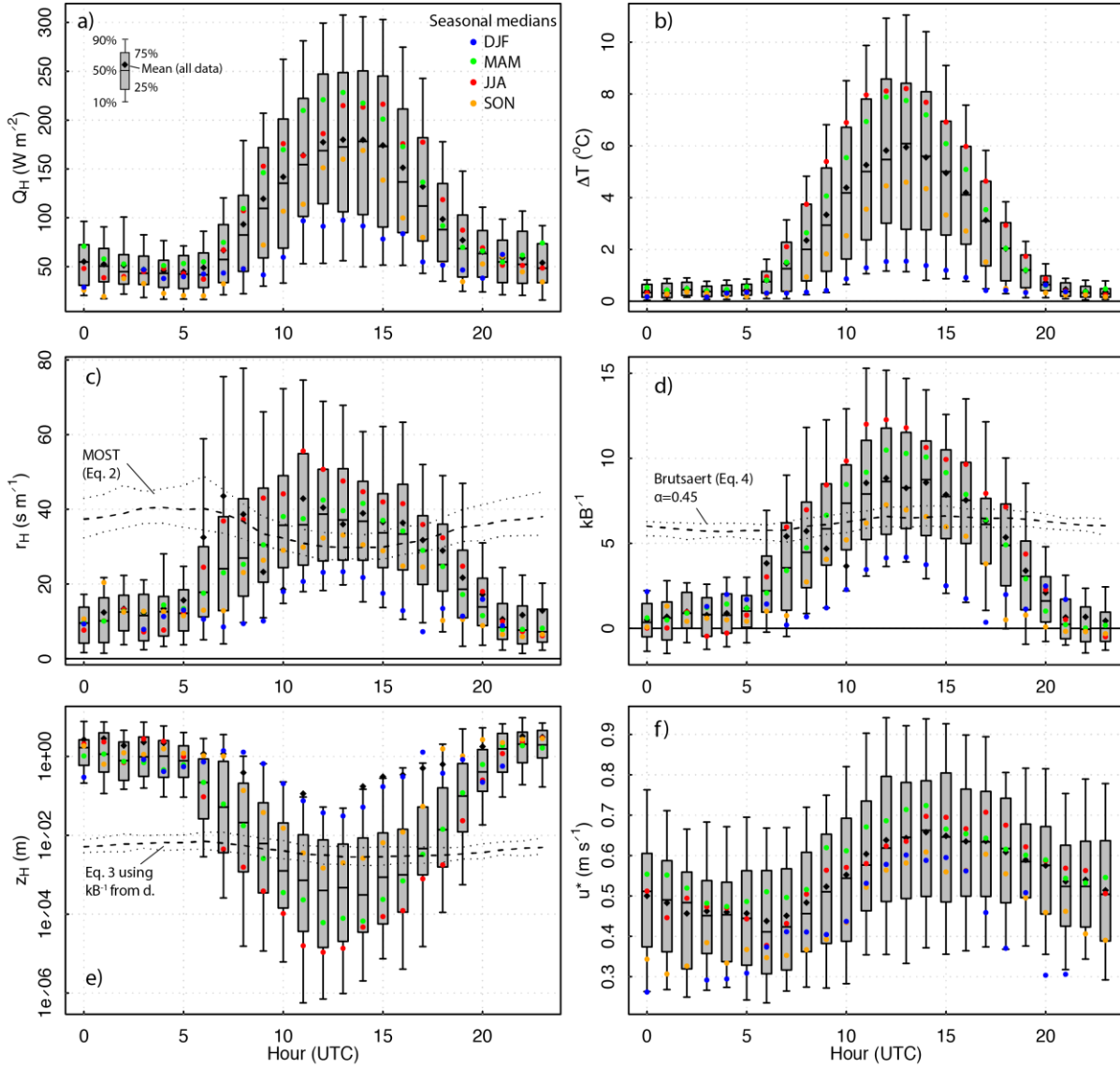


Figure 7. Diurnal course of hourly a) EC  $Q_H$ , b)  $T_R-T_A$ , c)  $r_H$ , d)  $kB^{-1}$ , e)  $z_H$ , and f)  $u^*$  from KCL measurements during 2015 or calculated using Eq. 2-3. Seasonal (DJF, MAM, JJA, SON) medians (coloured dots) are shown within the boxplot which are based on all 2015 data. Note e) y-axis scale is logarithmic. For c-e), boxplots and dots are derived from KCL measurements and are compared to hourly median (dashed line) and IQR (dotted line) results from the other methods (Appendix 1 Figure A1, Table A2).

### 3.3 Temporal variability of $z_H$

Median hourly  $z_H$  values are small throughout the day but vary across several orders of magnitude (Figures 7e, 8). Here,  $z_H$  is calculated using Eq. 3, based on  $kB^{-1}$  determined from observations (Eq. 2, using  $r_H$  calculated from measurements [Eq. 1]), with  $z_0$  calculated for  $10^\circ$  wind sectors using the Kanda et al. (2013) method (Section 2). The diurnal cycle of  $z_H$  in Figure 7e generally agrees with that calculated using Eqs. 3 and 4 (though this range of values is much smaller). Highest values are found overnight (up to 2.9 m at 00:00 h) and lowest during midday ( $1.3 \times 10^{-4}$  m at 12:00 h).

In central London,  $z_H$  is found to vary with solar angle above the horizon, consistent with results from Vancouver (Voogt and Grimmond 2000), but in contrast to findings over the COSMO scale model (Kanda et al. 2007). There is a possibility of a spurious correlation between solar angle and  $z_H$  due to turbulence, which also is expected to vary with solar angle. However, we find similar patterns during both neutral ( $-0.1 < z'/L < 0.1$ ) and unstable ( $z'/L < -0.1$ ) conditions. This relation is the basis for method R2 (Section 2.5).

We also find  $z_H$  varies with wind speed (R3, Section 2.5), turbulence (standard deviation of 10 Hz vertical wind speed,  $\sigma_w$ , measured by the KCL sonic anemometer), and stability ( $z'/L$ ). As mean wind speed and  $\sigma_w$  increase,  $z_H$  tends to decrease and median  $z_H$  is lower (higher) during unstable (neutral) conditions (Figure 8). Here, only daytime ( $K \downarrow > 10 \text{ W m}^{-2}$ ), non-stable ( $Q_H > 20 \text{ W m}^{-2}$ ), dry surface conditions (time since rain  $> 12$  h) are analysed to avoid uncertainties during overnight, reduced mixing periods.

These observations can be explained conceptually by analogy with a surface boundary layer. By definition,  $z_H$  is the height at which  $T_0 = T_A$ . When there are unstable conditions and vigorous turbulence, cool air is mixed downwards into the thin layer between the surface and  $z_H$ . This acts to lower  $z_H$ , and overall heat transfer is enhanced by turbulent transport



above  $z_H$ . In contrast, during neutral or stable conditions with minimal turbulence, the  $z_H$  height grows as  $Q_H$  from the surface acts to warm the air in contact with the surface via molecular diffusion.

This is somewhat contradicted however by the observed relation between increasing surface bulk heat transfer coefficient ( $C_H$ ) and increasing  $z_H$ , calculated (Kanda et al. 2007):

$$C_H = \frac{Q_H}{\rho c_p U (T_R - T_A)} = \frac{1}{r_h U} \quad (12)$$

A possible explanation for this trend is related to the different source areas influencing radiometers and turbulent EC measurements (as indicated for  $r_h$  in Figure 7).

Overall, it remains difficult to interpret the large diurnal variations in  $z_H$  and their physical basis. These variations could be an artefact of the measurements used to derive  $z_H$ . For example, the apparent relation between  $z_H$  and solar angle may be due to differences in  $LST$  and the effective surface temperature ( $T_E$ ). According to Kanda (2005),  $T_E$  is the operative driver of  $Q_H$  and is formulated as the average of individual facets' surface temperatures weighted by a local bulk transfer coefficient. Differences between  $T_E$  and practical substitutes such as  $T_R$  or  $LST$  used in Eq. 1 vary diurnally according to solar angle and resulting thermal anisotropy of the 3d urban surface and are thus manifested in derived  $z_H$ . In general, targeted, controlled experiments may offer a more suitable approach for more in-depth investigation of  $z_H$ . However, this requires more detailed surface temperature measurements than can be obtained currently by EO and is the driver behind other ongoing work (e.g. Morrison et al. 2018).

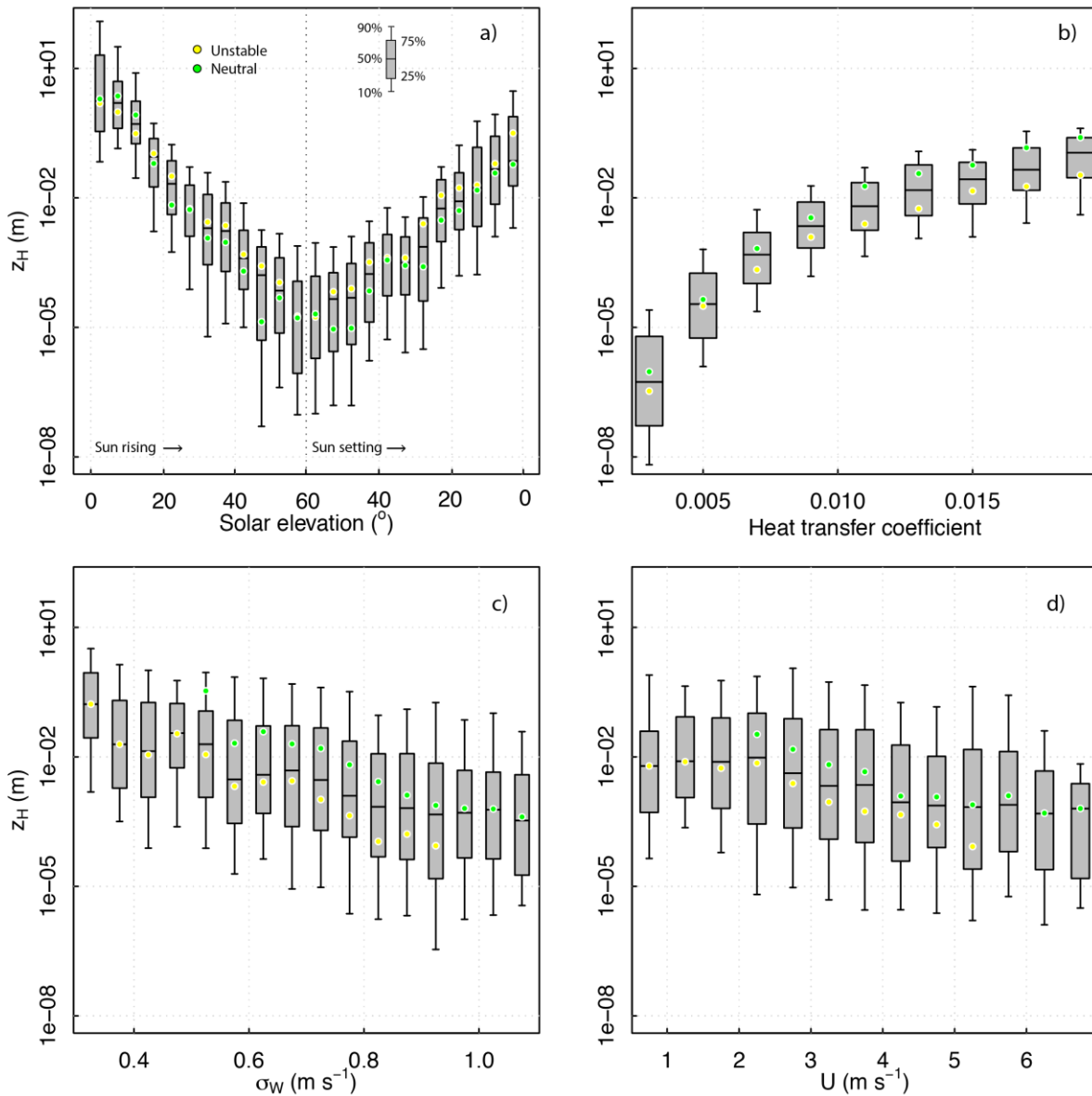


Figure 4. Normalized  $LST$  for all available Landsat images binned by a) building land cover fraction with medians for meteorological season (coloured dots) and b) mean building height with medians for building land cover fraction (coloured dots). In a) only pixels with water < 0.5 are used and in b) only pixels with vegetation and water < 0.5 are selected.

### 3.4 Influence of $Q_F$

Another factor affecting *in situ* observations with implications for the ARM method is the influence of  $Q_F$ . It is well-known that scintillometers and EC  $Q_H$  observations in cities also include a portion of  $Q_F$  (e.g. Crawford et al. 2017). Modelling work in central London suggests that between 2/3 (Bohnenstengel et al. 2014) and 80% (Iamarino et al. 2012) of  $Q_F$  is expressed as  $Q_H$ , with the remainder partitioned towards  $L\uparrow$ , latent heat flux ( $Q_E$ ), storage heat flux ( $\Delta Q_S$ ), and waste water. This contribution will vary by time of day, day of week, and season according to local energy consumption for space heating, vehicle transportation, and lighting, *etc.*

In terms of the ARM approach, the overall effect of  $Q_F$  is complex. Direct  $Q_F$  emissions to the atmosphere (e.g. from vehicle exhaust or building chimneys) will warm the atmosphere as a function of the urban boundary layer (UBL) height (with dynamic feedback) and  $Q_F$  input. A simple calculation assuming a well-mixed UBL shows that for a constant  $100 \text{ W m}^{-2}$   $Q_F$  directly emitted to the atmosphere over 30 minutes, warming would range from 0.10-0.75 K for UBL heights of 1000 m and 200 m, respectively. This agrees with modelling work that estimates an increase of up to 1.5 K in screen-level air temperature from  $Q_F$  during winter nights (Bohnenstengel et al. 2014).

In an ARM context, an increase in daytime air temperature alone would *decrease*  $Q_H$  by reducing the vertical gradient between  $T_0$  and  $T_A$ . We would expect some corresponding increase in  $T_0$  from  $Q_F$  to counteract this, but the location, timing, and magnitude of the surface warming is less well understood. Bohnenstengel et al. (2014) show that approximately 1/3 of  $Q_F$  goes into surface heating (and thus outgoing longwave radiation), but that there is a delay of several hours due to the thermal mass of the urban fabric and diffusion of indoor heating sources outwards to exterior building surfaces. Where this warming occurs also has implications for EO-observed *LST*. If building warming is preferentially directed towards external vertical walls rather than roofs, this may not be captured by downward facing EO sensors with a view-bias towards horizontal rooftops.

To assess whether there are thermal differences in an urban setting attributable to  $Q_F$ , observations from the flux tower at KCL in central London data are sorted into working (M-F) and non-working days (S-S and public holidays). Friday evenings (after 19:00 h) are defined as ‘non-working’ and Sunday evenings (after 19:00 h) as ‘working’ (Björkegren and Grimmond 2016). Three winter seasons (DJF 2013-14, 2014-15, 2015-16) are used for analysis because  $Q_F$  inputs are expected to be highest relative to  $Q^*$  and more easily detected than during other seasons. The 30-minute analysis data are also controlled for  $K\downarrow$  (so that hourly median working and non-working day’s  $K\downarrow$  values are all within  $\pm 5 \text{ W m}^{-2}$ ) to minimize differences in surface solar energy input.

This analysis shows small but detectable differences in mean hourly air temperature and surface temperature between working and non-working days (Figure 9). Working weekdays tend to have slightly warmer  $T_R$  (up to 0.7 K) and air temperatures (up to 0.5 K) compared to weekends and holidays. Working days are generally warmer during daytime with largest differences at midday (1100-1200 UTC). Overnight, weekends and holidays are warmer than working days.

This translates to  $T_R$ - $T_A$  differences between working and non-working days as well, with highest differences up to 0.4 K (up to 66%). This vertical gradient equates to  $Q_H$  differences of around  $20 \text{ W m}^{-2}$  during daytime, assuming an  $r_H$  value of  $25 \text{ s m}^{-1}$  representative of winter daytime conditions (Figure 7c). This is comparable to observed eddy covariance  $Q_H$  differences of  $4.2 - 20.6 \text{ W m}^{-2}$  (0900-1800 UTC).

The diurnal cycles of observed temperature differences are compared to the diurnal cycle of  $Q_F$  as modelled by the new version of the Greater $Q_F$  model (Iamarino et al. 2012): GQF (Gabey et al. 2018) within UMEP (Lindberg et al. 2018). The shape of the model diurnal  $Q_F$  cycle is based on energy consumption statistics and varies by season, day of week and combination of building sources within an area (Gabey et al. 2018); i.e. KCL tower measurement source areas. Energy consumption statistics demonstrate that  $Q_F$  is dominated by building sources and indoor space heating from natural gas combustion. To directly compare the shape of the diurnal course between observed temperature differences and modelled  $Q_F$ , both are plotted on the same figure with different y-axes (Figure 9).

From this comparison, we see that modelled  $Q_F$  differences follow a diurnal pattern with higher working day (non-working day) values during day (night). Modelled energy consumption differences peak at 09:00 UTC, while peak observed  $T_A$  differences occur at 11:00 UTC and  $T_R$  differences are at 12:00 UTC. This suggests a lag of 2-3 hours between energy consumption ( $Q_F$  emissions) and when external air and surface warming occurs.

There are several implications resulting from this lag. The first is that  $Q_F$  emissions are embedded in observed *LST* and  $T_A$ , and thus also  $Q_H$  when using ARM; though instantaneous *LST* (as observed from satellite-platforms) and  $T_A$  will not necessarily reflect the actual timing of  $Q_F$  emissions. The magnitude of the  $Q_F$  contribution to  $Q_H$  will depend on the city location and time of day and year and may often only be a minor energy source, especially relative to uncertainties in other aspects with determining  $Q_H$  with ARM (Section 3.3-3.4). In principle, this suggests using a residual approach to determine  $Q_F$  based on remote sensing is possible (e.g. Chrysoulakis et al. 2015), though current uncertainties in determining  $Q_H$  are on the same order of magnitude as the  $Q_F$  signal (Section 3.3-3.4).

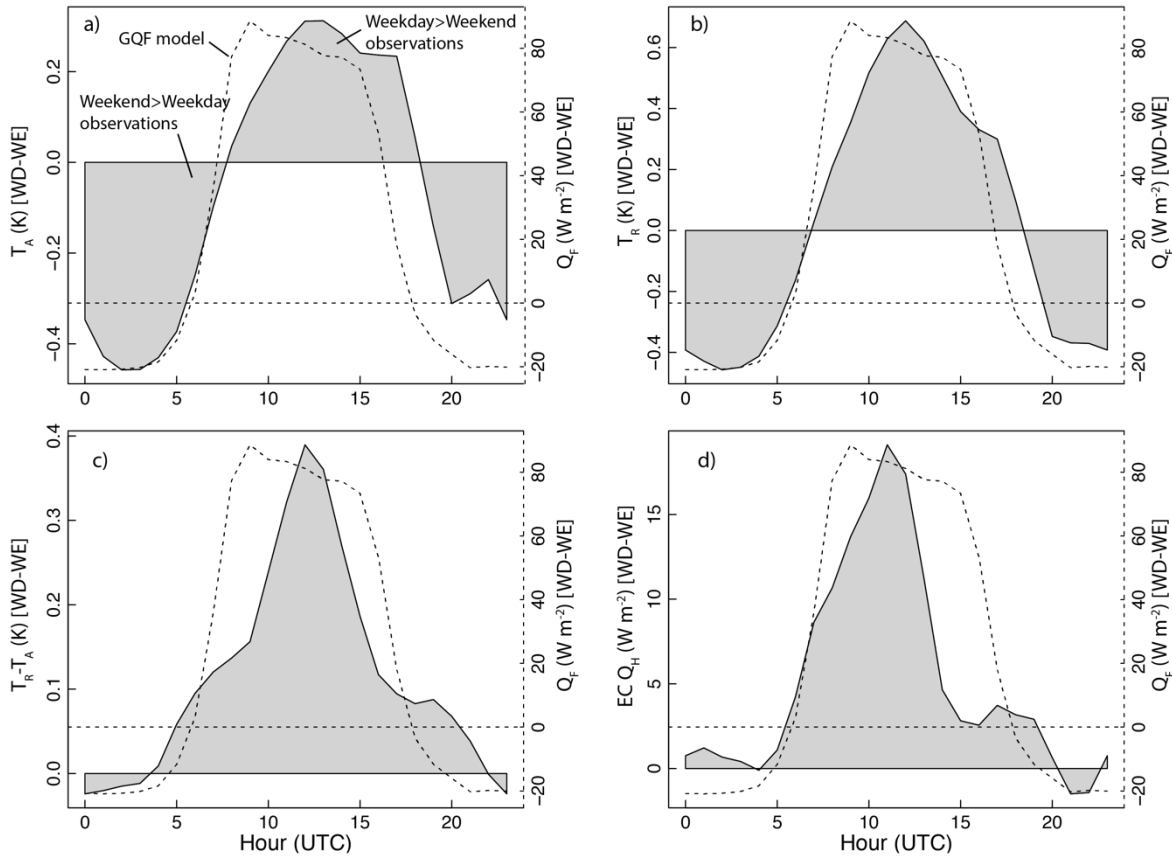


Figure 9. Working (weekday) and non-working (weekend) average differences for a)  $T_A$ , b)  $T_R$ , c)  $T_R - T_A$ , and d) eddy covariance  $Q_H$ . All data are from the KCL flux tower for 2013–2015 DJF seasons. Results from the energy inventory-based GQF (Iamarino et al. 2012, Gabey et al. 2018, Lindberg et al. 2018) model are shown for reference using second y-axis (dashed line). Note the temperature y-axis ranges for (a)–(c) differ.

### 3.5 $Q_H$ using ARM

Domain  $Q_H$  is calculated with an aerodynamic resistance method (ARM) using observed  $LST$ ,  $T_A$ , and calculated  $r_H$  (Section 2.5). A single  $LST$  scene (2 Oct 2015, 11:00 UTC) is used as a case study to exemplify results and allow a more detailed spatial view (Figure 10). This period is selected because of high image coverage (97% cloud free) during the satellite overpass and availability of observations from the measurement network for comparison. The day was a working weekday (Friday) with clear skies and mild temperatures (observed 15.7 °C at KCL tower at 11:00 UTC). There was a small amount of precipitation (0.04 mm) recorded the previous day at 15:00 h, but the surface is assumed to have dried by the time of the overpass because this occurred more than 12 hours earlier (Kotthaus and Grimmond 2014). Winds were calm (2.6 m s<sup>-1</sup>) and from the NE (78°), typical of high-pressure anti-cyclonic synoptic conditions. Sunrise was at 6:02 UTC and solar elevation at the time of the overpass was approximately 30° above the horizon. The previous night was generally clear and calm, allowing for enhanced radiative cooling of the surface overnight.

At the 11:00 UTC overpass, EO-derived  $LST$  ranged from 14–28 °C, with cooler temperatures measured generally in vegetated park areas, water bodies, and city centre areas with tallest buildings.  $T_A$  varies within a much smaller range (15.6–16.2 °C) compared to  $LST$  with cooler values in the city centre associated with higher building heights and thus higher reference height.

Spatial patterns of  $LST$  (and input meteorological values) are also manifested in  $r_H$  and  $Q_H$ . Excluding parks and water bodies, all  $Q_H$  methods (R1–R5) show similar spatial pattern of lower values in the city centre and highest values in more vegetated, less densely built areas. This spatial pattern is somewhat counterintuitive as it was expected that highest  $Q_H$  would be found in areas with highest built densities with associated higher Bowen ratios and higher  $Q_F$  releases compared to less densely built and populated areas. This primarily results from observed  $LST$  patterns in central London (Figures 3–6).

Although spatial patterns are similar, there are notably different distributions of pixel-scale  $Q_H$  values between the methods. R5 tends to predict highest values compared to the other versions for individual pixels away from the city centre. These values (30% >500 W m<sup>-2</sup> for all 25 scenes) are associated with higher vegetation land cover fractions that result in lower  $\alpha$ , lower  $kB^{-1}$ , and lower  $r_H$ . Whereas,  $Q_H$  values determined using R4 appear low for an urban area (25% from 50–100 W m<sup>-2</sup>). This is not unexpected given differences in form (vegetation) and function ( $Q_F$ ) between the physical model on which R4 is based and the study area, and in Section 3.6 we take advantage of this difference to constrain  $Q_F$  estimates.

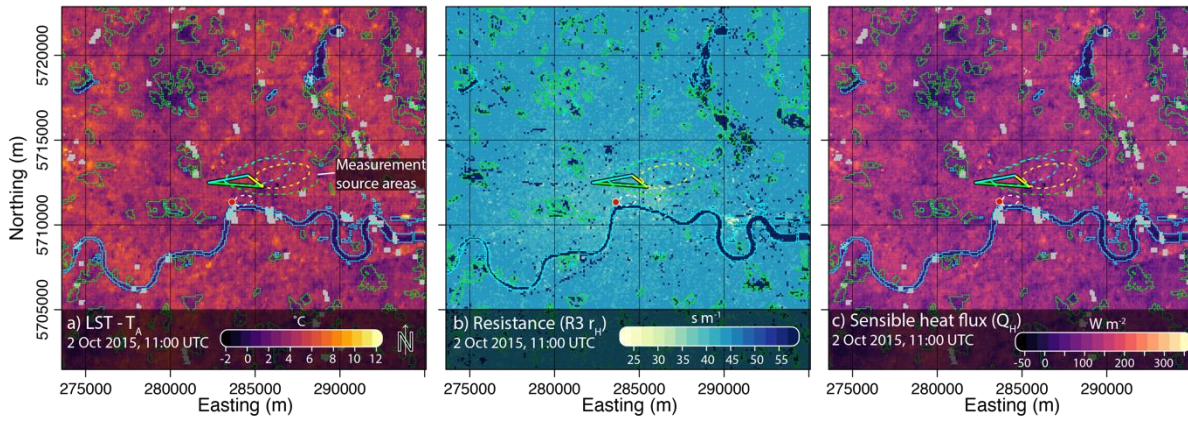


Figure 10. a)  $LST-T_A$ , b)  $r_H$ , and c)  $Q_H$  for 11:00 UTC, 2 October 2015 for the R3 method (Table 2). R3 is used as representative of R1-R3 methods (techniques based on central London flux tower observations). Other symbols and outlines are as in Figure 1. Cloud pixels are masked (grey).

To quantify sensitivity of  $Q_H$  results to individual input variables for each method,  $Q_H$  is modelled with each input variable perturbed individually, while all other variables remain constant (Figure 11). Each variable is perturbed by both a constant  $\pm 10\%$  and the uncertainty ranges listed in Table 3. The case study date (2 Oct 2015) is analysed as representative.

Unsurprisingly, all methods are most sensitive to perturbations in  $LST$  ( $\pm 50$ -70%) and to a lesser extent  $T_A$  (40-60%) (Figure 11).  $Q_H$  is also very sensitive to  $u^*$  for methods R2, R4, and R5, as this is used to determine  $kB^{-1}$  and gridded wind speed. The R3 method, which uses wind speed as a primary input, is also moderately sensitive to  $z_0$  and  $z_d$ . The R5 method is very sensitive to  $\alpha$  and  $\lambda_V$  as this determines  $kB^{-1}$  and  $r_H$ . Results are relatively insensitive to  $L$  and surface variables  $z$  and  $z_d$ .

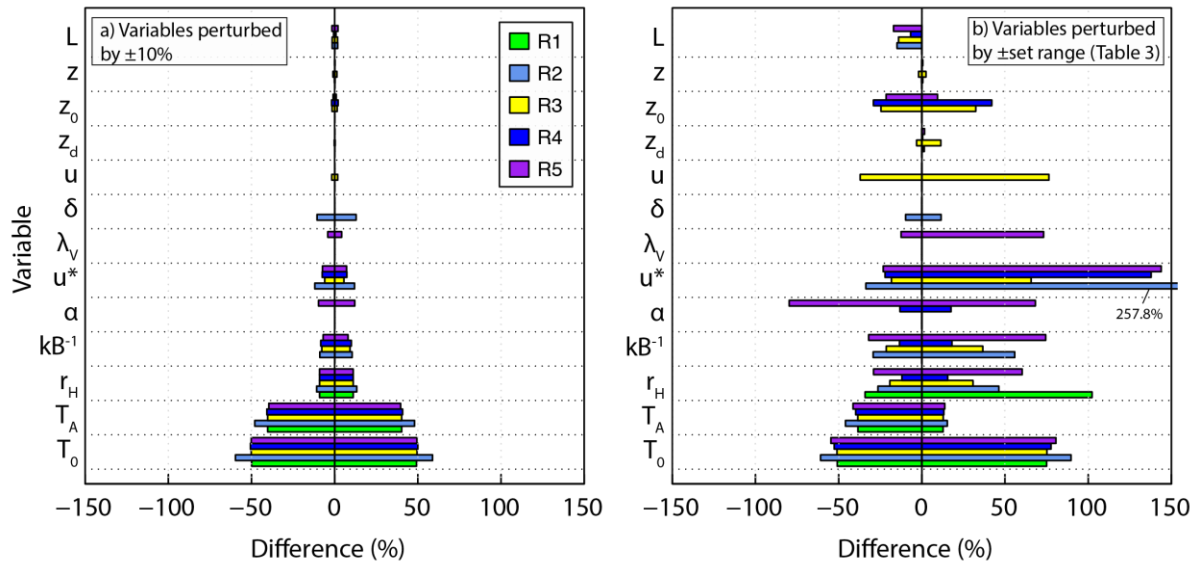


Figure 11. Percentage difference in  $Q_H$  for five methods (Tables 2, A2) with individually perturbed input variables and parameters relative to a control run. Each input is perturbed by a)  $\pm 10\%$  and by b) prescribed ranges given in Table 3. Also see Table 3 for definition of variables. Note that some methods do not require all variables (Table A.2).

### 3.6 Comparison to measurements

Each method (R1-R5, Section 2) is compared to observations, to assess the uncertainties and associated assumptions of each. The goal is to understand the urban system based on the entire suite of models, rather than to assign a relative ranking of the models.

Overall, when  $Q_H$  is modelled using five methods for all 25 available images, there is an expected annual cycle with highest values generally in the spring ( $\sim 300$ -400  $W m^{-2}$ , April-May) and lowest values in the autumn and winter ( $< 150$   $W m^{-2}$ , October-February) (Figure 3b). Individual R1-R5 results are compared to observations from the three scintillometers and EC station using turbulent flux source area models which vary by observed wind direction and atmospheric conditions ( $u^*$ ,  $L$ ,  $\sigma_V$ ) at KCL. The source area model of Kormann and Meixner (2001) is used and full details for implementation and weighting along the scintillometer path are described in Crawford et al. (2017). The source area weights for each instrument (100 m resolution) are multiplied by the 100 m  $Q_H$  pixels to determine the source area weighted average  $Q_H$  to compare with observations.

Overall, R1-R3 (i.e. based on London observations)  $Q_H$  compares expectedly reasonably well to observations, though there is scatter about the 1:1 line (Figure 12). For all observations, the lowest mean absolute error (MAE) relative to observations is for R1 (58.1 W m<sup>-2</sup>) (Table 4). The R4 method (Kanda et al. 2007) tends to underestimate observed  $Q_H$  by 129.3 W m<sup>-2</sup>, on average (Table 4). This too is somewhat expected given differences between the model and real city. We interpret this result as a useful reference or baseline representing  $Q_H$  from a non-vegetated city with no active  $Q_F$  sources. From this perspective, we can attempt to attribute and constrain contributions of additional energy sources in a real city.

There are likely two primary reasons for the R4-observed difference: 1) vegetation in the real city which acts to lower  $kB^{-1}$  and  $r_H$  and thus increase  $Q_H$ , and 2)  $Q_F$  which is embedded in  $Q_H$  observations from live cities and therefore also the empirical models based on these observations. Additional physical differences between the model and real world (i.e. building orientation, spacing, and alignment) are less likely to have an appreciable effect. This is based on COSMO model tests with different configurations that demonstrate R4 is applicable over a range of building geometries and alignments (Kanda and Moriizumi, 2009). For vegetation,  $\lambda_v$  in central London is relatively sparse (~0.09) and actual  $\alpha$  and  $kB^{-1}$  values are lower than expected based purely on this vegetation fraction (Figure 2b).

The primary reason for the discrepancy then may be anthropogenic heat flux ( $Q_F$ ). The R4 method is based on EC observations from a scale model city with no  $Q_F$  sources, whereas the other methods are based on EC observations from real cities which include a portion of  $Q_F$  embedded in  $Q_H$  measurements. Thus, the R4 method may account for the form but not the functioning of a city and thus underestimates  $Q_H$  measurements which include a portion of  $Q_F$  in real cities.

To assess if the difference between R4 results and observations could be used to constrain  $Q_F$  estimates, total  $Q_F$  values were calculated for a subset of overpass times using an independent energy model for spatial areas within the measurement source areas (Gabey et al. 2018). Approximately 80% of total  $Q_F$  is expected to be manifest as  $Q_H$ , (as noted in Section 3.4;  $L\uparrow$ ,  $Q_E$ ,  $\Delta Q_S$ , and waste water also account for the remainder). The GQF model  $Q_F$  expressed as  $Q_H$  ranges from 111 – 180 W m<sup>-2</sup> and agrees well with the 129.3 W m<sup>-2</sup> MAE between the R4  $Q_H$  model and observations. The same approach also works using the difference between modelled R1  $Q_H$  (constant  $r_H$ ) and R4 (123.3 W m<sup>-2</sup>). Thus, it appears that the R4 method calculates non-anthropogenic  $Q_H$  with no added  $Q_F$  component and may have potential as another tool to constrain  $Q_F$ , in combination with  $Q_H$  observations.

The difference in magnitude of  $Q_H$  calculated using winter-observed  $r_H$  and summer-observed  $r_H$  may also be useful to provide a constraint on  $Q_F$ . This is tested with a simple calculation assuming constant ( $T_R - T_A$ ) of 4 °C and typical mid-day winter and summer values of  $r_H$  (20 and 50 s m<sup>-1</sup>, respectively: Figure 7c). This leads to a difference in mid-day  $Q_H$  of 144 W m<sup>-2</sup>, similar in magnitude to modelled GQF  $Q_F$ . The result implies a more efficient heat transfer in winter, but it is unclear if this is from increased  $Q_F$  magnitude (but without a corresponding observed change in  $T_R - T_A$ ) or from seasonal differences in turbulence regimes and surface conditions.

Though there is scatter in the comparisons (coefficient of determination  $R^2$  range from 0.42 [R5] to 0.59 [R4]), 80% of all values fall within the 80% uncertainty range (Figure 12) determined from the Monte Carlo sampling technique (Section 2). This implies modelling results are within realistic range of observations, given limitations of the accuracy and resolution of input variables.

Table 4. Observed and source area-averaged model (R1-R5) comparison mean absolute error (MAE) for  $Q_H$ . MAE values (W m<sup>-2</sup>) are calculated relative to observations for each sensor and all sensors combined. N = 25 images/time periods.

	P1	P2	P3	EC	All
R1	35.1	56.2	34.2	75.6	58.1
R2	81.6	22.8	45.5	82.9	66.8
R3	47.4	53.2	68.6	73.7	64.5
R4	166.9	97.9	112.8	130.6	129.3
R5	52.8	75.6	62.5	83.6	73.0
R4-5 avg.	33.9	77.9	54.8	64.5	59.3

Based on Monte Carlo analysis of input variables (Table 5), R5 has a broad distribution of values (i.e. more uncertain prediction) because the method relies on several input variables with relatively high degrees of associated uncertainty, particularly  $u^*$ ,  $\alpha$ , and vegetation land cover fraction (Table 3, Section 3.5). R3 has relatively high overall uncertainty due to uncertainties in calculating spatial wind speed variations, which also depend on  $u^*$  and  $z_0$ . Probability distributions of R1 and R4 are relatively narrow (i.e. more precise prediction, though not necessarily more accurate) due to reliance on fewer input variables.

Overall, relative uncertainty tends to increase with vegetation land cover fraction ( $\lambda_v$ ), particularly for R5 (Figure 13). This is due to the dependence of this method on  $\lambda_v$  to determine  $\alpha$ ,  $kB^{-1}$ , and  $r_H$ . Correlation with land cover is less evident with other methods, though the overall uncertainty with R3 is consistently high and R2 and R4 are relatively low. Based on this analysis, we conclude the Kanda et al. (2007) method (R4) performs well at  $Q_H$  predictions based on physical form of the city and using an aerodynamic approach. It is of a generalized form for use in a range of cities and is relatively insensitive to variations in meteorological and morphological inputs. However, it is important to note that it is calibrated to a scale model city without  $Q_F$  sources (nor vegetation), so predictions are likely of  $Q_H$  without additional  $Q_F$  inputs found in real cities. Thus, it tends to underestimate observed  $Q_H$  in cities which also include a portion of imbedded  $Q_F$ . From this perspective, the method also has potential to serve as a baseline reference to assess additional sources of urban energy exchange.

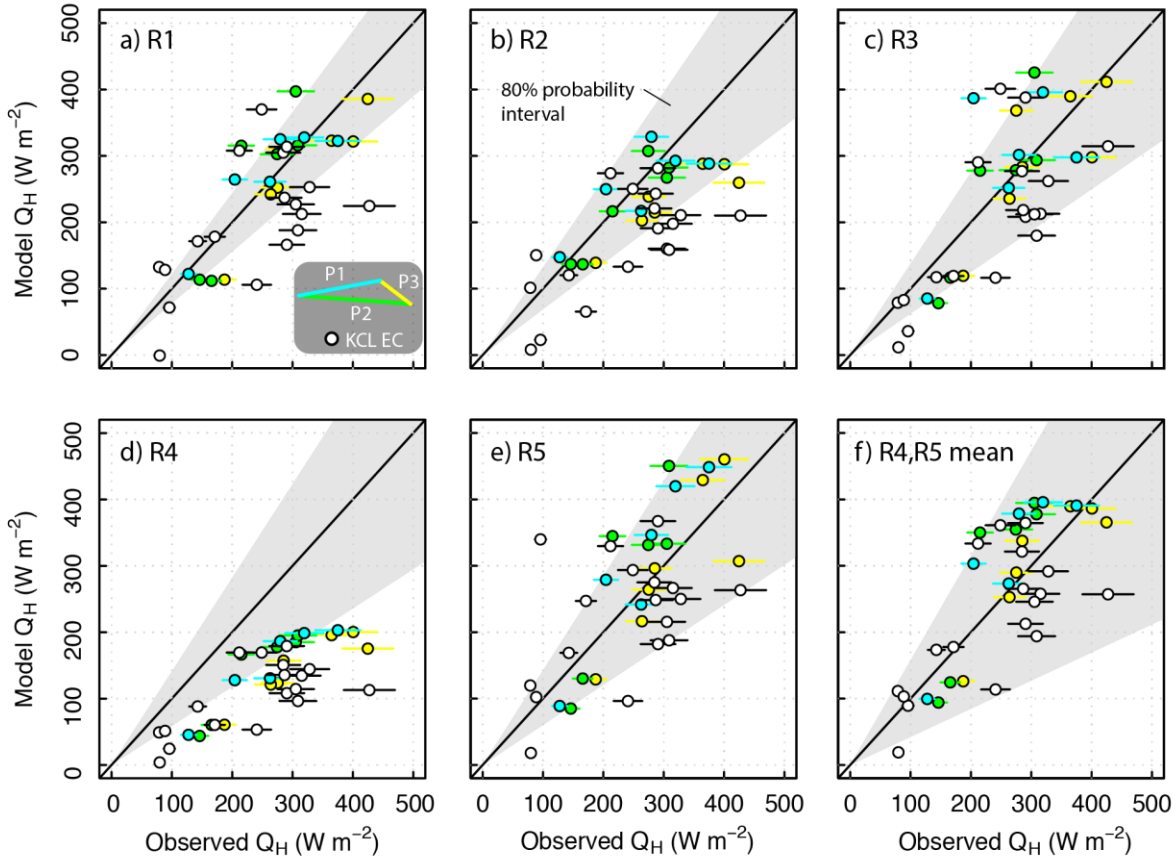


Figure 12. Source area-averaged model  $Q_H$  and observations from the scintillometer network (coloured points) and EC tower (white points) with 80% model probability intervals generated for each method (Table 3) from Monte Carlo analysis (grey shaded area). Horizontal error bars are measurement uncertainties (Section 2.2). Comparison statistics are given in Table 4.

Table 5. Absolute and relative  $Q_H$  uncertainties for each method (R1-R5, Table 2) for the entire model domain based on Monte Carlo analysis. These values are the mean range of 80% of values for each pixel for each scene based on 20 model runs with varied input parameters.

	R1	R2	R3	R4	R5
Absolute ( $W m^{-2}$ )	52.8	56.7	78.6	28.2	27.7
Relative (%)	26	46	68	48	51

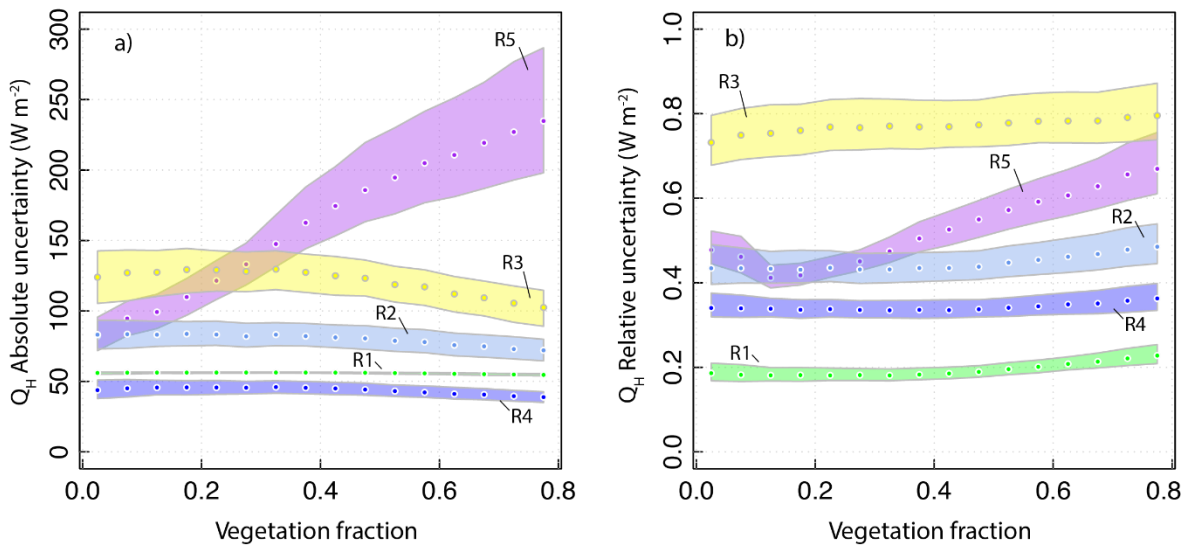


Figure 13. Absolute and relative uncertainty (probability) intervals for ARM-modelled  $Q_H$ . The probability interval encloses 80% of  $Q_H$  values generated for each pixel based on Monte Carlo analysis for all 25 images. Relative probability interval is defined as the absolute 80% interval range divided by the median for each pixel. Pixels are sorted by vegetation fraction (0.05 bin width) and the interquartile range (shaded area) and median (points) of probability intervals are shown for each resistance method (R1-R5).



## 4 Conclusions

In this work, we analyse an EO-derived *LST* dataset spanning six years to assess the challenges of using an aerodynamic resistance approach to generate high spatial resolution (100 m) estimates of urban  $Q_H$ . A unique feature of this study is the presence of a  $Q_H$  measurement network of scintillometers and eddy covariance with which to compare modelled  $Q_H$ .

*LST* spatial patterns are linked with the distribution of vegetation and buildings at the surface. Areas with tall buildings are cooler than average (up to 10%) in *LST* images, but it remains to be seen if this is due to increased shading and reduced sky-view factors or from uncertainties with surface bulk surface emissivity values or satellite view angle bias and thermal anisotropy.

Analysis of  $T_R$  and  $T_A$  also demonstrates an anthropogenic influence with implications for ARM  $Q_H$ . Working weekday  $T_R$  and  $T_A$  are warmer than weekends and holidays during daytime and there is a lag of 2-3 hours between  $Q_F$  emissions (based on energy consumption inventory models) and observed external surface and air warming. This also highlights a shortcoming of using EO-*LST* data: it provides a high spatial resolution snapshot of conditions at a single instant but only provides low temporal resolution.

To determine  $Q_H$  from *LST*, five thermal resistance methods are compared. Three (R1-R3) are based on empirical relations observed at a central London EC tower: i) observed diurnal cycle of  $r_H$ , ii) a linear relation between measured wind speed and  $z_H$ , and iii) a linear relation between solar elevation and  $z_H$ .

Analysis to derive these relations demonstrate that  $r_H$ ,  $kB^{-1}$ , and  $z_H$  vary on diurnal and seasonal timescales. Diurnally,  $r_H$  and  $kB^{-1}$  are greatest, and  $z_H$  is least, during mid-day periods. This is reasoned to be because momentum transfer is enhanced relative to heat transfer in urban areas due to pressure drag from bluff bodies. Although  $Q_H$  is also greatest during mid-day periods, the momentum transfer is comparatively greater. In general, observations of  $z_H$  in urban areas are rare and its variability is poorly understood. Here we show  $z_H$  varies diurnally according to wind speed, turbulence, solar angle and atmospheric stability.

Observations and models of  $Q_H$  using these methods are in general agreement, though there is scatter in the linear comparisons. Model results are found to be sensitive to uncertainties in  $T_0$  and input meteorological variables needed to estimate  $r_H$  ( $u$ ,  $u^*$ ,  $T_A$ ). However, based on Monte Carlo sampling of these variables within prescribed ranges, 80% of comparisons are within 80% confidence intervals.

Two previously published empirical relations are also tested, derived from observations based on: i) a physical scale urban model (R4) (Kanda et al. 2007) and ii) from several real cities (R5) (Kawai et al. 2009). For the R4 method, differences between observations are most likely due to the influence of  $Q_F$  on observations. Since the model coefficients are calibrated to a scale model with no  $Q_F$  sources, results using this method may underestimate  $Q_H$  observations in a real city, which include a portion of  $Q_F$ . Thus, this method may have potential to act as a baseline reference and constrain estimates of energy partitioning between  $Q_H$  and  $Q_F$ .

For the R5 method, pixels with greater vegetation land cover tend to produce unrealistically high  $Q_H$  values. This is largely due to uncertainties in determining  $kB^{-1}$  parameters across the full spectrum of urban forms and neighbourhood types.

In general, R4 (Kanda et al. 2007) is suitable to calculate  $r_H$  because of its generalized urban form and high precision due to reliance on relatively few input variables. This work also highlights the need for further measurements in contrasting urban environments: particularly in areas with tall, dense buildings and neighbourhoods with high vegetation density. Ideally a range of observations that include a range of urban form (e.g. denser/lighter building materials; smaller/larger buildings) and function (e.g. more/less  $Q_F$ ) with changing wind flow regimes (i.e. roughness element packing) would allow a systematic assessment of the controls on  $r_H$  prior to adding the vegetation effects.

Overall, an aerodynamic resistance approach remains a useful, conceptually straight-forward way to estimate bulk  $Q_H$  in urban areas. However, the approach is sensitive to several input variables (e.g.  $u^*$ ), so practical challenges remain in resolving spatial variations in urban surface and meteorological conditions in urban areas. Thus, in future studies a more probabilistic approach to modelling urban energy fluxes may be advantageous in the face of these uncertainties.

## 5 Acknowledgments

This work is funded by European Union Horizon 2020 UrbanFluxes, the Newton Fund/Met Office CSSP-China, NERC ClearfLo, NERC TRUC, and NERC AirPro. We are very grateful to BT, Islington Council, Barbican, and King's College London for site access. Kjell zum Berge, Ting Sun, William Morrison, Natalie Theeuwes, and Simone Kotthaus (University of Reading) and Makoto Nakayoshi (Tokyo University of Science) made valuable contributions to fieldwork and discussions.

## 6 References

- Arnfield, A. J. "An approach to the estimation of the surface radiative properties and radiation budgets of cities." *Physical Geography* 3.2 (1982): 97-122.
- Barlow J, Best M, Bohnenstengel SI, Clark P, Grimmond S, Lean H, Christen A, Emeis, S, Haeffelin M, Harman IN, Lemonsu A, Martilli A, Pardyjak E, Rotach MW, Ballard S, Boutle I, Brown A, Cai X, Carpentieri M, Coccal O, Crawford B, Di Sabatino S, Dou J, Drew D, Edwards JM, Fallmann J, Fortuniak K, Gornall J, Gronemeier T, Halios C, Hertwig D, Hirano K, Holtslag AAM, Luo Z, Mills G, Nakayoshi M, Pain K, Schlünzen KH, Smith S, Soulhac L, Steeneveld G-J Sun T, Theeuwes N, Thomson D, Voogt JA, Ward H, Xie Z-T, Zhong J 2017: The integration of urban atmospheric processes across scales. *BAMS*, **98**, ES261–ES264 [doi:10.1175/BAMS-D-17-0106.1](https://doi.org/10.1175/BAMS-D-17-0106.1)



- Best, M. J., and C. S. B. Grimmond. "Key conclusions of the first international urban land surface model comparison project." *Bulletin of the American Meteorological Society* 96.5 (2015): 805-819.
- Brutsaert, W. H. "Evaporation in the Atmosphere." D. Reidel (1982).
- Chen, Yingying, et al. "Improving the Noah land surface model in arid regions with an appropriate parameterization of the thermal roughness length." *Journal of Hydrometeorology* 11.4 (2010): 995-1006.
- Christen, Andreas, and Roland Vogt. "Energy and radiation balance of a central European city." *International journal of climatology* 24.11 (2004): 1395-1421.
- Chrysoulakis, N., Heldens, W., Gastellu-Etchegorry, J.P., Grimmond, S., Feigenwinter, C., Lindberg, F., Del Frate, F., Klostermann, J., Mitraka, Z., Esch, T. and Al Bitar, A., 2015. Urban Energy Budget Estimation from Sentinels: The URBANFLUXES Project. *Mapping Urban Areas from Space*.
- Crawford B, CSB Grimmond, HC Ward, W Morrison, S. Kotthaus 2017: Spatial and temporal patterns of surface-atmosphere energy exchange in a dense urban environment using scintillometry *QJRMS*, 143, 703, B, 817–833 [10.1002/qj.2967](https://doi.org/10.1002/qj.2967)
- Crawford, B., and A. Christen. "Spatial source attribution of measured urban eddy covariance CO2 fluxes." *Theoretical and Applied Climatology* 119.3-4 (2015): 733-755.
- Crawford, B, C. S. B. Grimmond, and A Christen. "Five years of carbon dioxide fluxes measurements in a highly vegetated suburban area." *Atmospheric Environment* 45.4 (2011): 896-905.
- Del Frate, Fabio, W. Heldens, N. Chrysoulakis, F. Lindberg, Z. Mitraka, C.S.B. Grimmond, A. Gabey, F. Olofson, D. Poursanidis, C. Feigenwinter, A. Albitar, D. Latini. Urban surface morphology land cover/use and characteristics. UrbanFluxes Report D3.1 (2016). DLR 2012, [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor3\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor3_manual_2012.pdf) Accessed 24 July 2016
- Gabey A, S Grimmond, I Capel-Timms (2018) Anthropogenic Heat Flux: At what spatial resolution is a simple model reasonable? *Theoretical and Applied Climatology* doi: 10.1007/s00704-018-2367-y
- Gabey A., S. Grimmond. 'Availability, evaluation and synthesis of metadata describing Citizen Weather Stations', In preparation.
- Gerace, A., & Montanaro, M. (2017). Derivation and validation of the stray light correction algorithm for the thermal infrared sensor onboard Landsat 8. *Remote Sensing of Environment*, 191, 246-257.
- GLA Datastore (2017), accessed 19 May, 2017. <https://data.london.gov.uk/dataset/interim-2015-based-population-projections/resource/4fcb8bfa-f860-48cc-af96-44a0931e4f27#>
- Grimmond, CSB, and TR Oke. "Aerodynamic properties of urban areas derived from analysis of surface form." *Journal of applied meteorology* 38.9 (1999): 1262-1292.
- Grimmond, C. S. B., M. Blackett, M. J. Best, J-J. Baik, S. E. Belcher, J. Beringer, S. I. Bohnenstengel et al. "Initial results from Phase 2 of the international urban energy balance model comparison." *International Journal of Climatology* 31, no. 2 (2011): 244-272.
- Heidbach, Katja, Hans Peter Schmid, and Matthias Mauder. "Experimental evaluation of flux footprint models." *Agricultural and Forest Meteorology* 246 (2017): 142-153.
- Högström, U. L. F. "Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation." *Topics in Micrometeorology. A Festschrift for Arch Dyer*. Springer Netherlands, 1988. 55-78.
- Iamarino, M., S. Beevers and C. S. B. Grimmond (2012). "High-resolution (space, time) anthropogenic heat emissions: London 1970–2025." *International Journal of Climatology* 32(11): 1754-1767.
- Kato, S., & Yamaguchi, Y. (2005). Analysis of urban heat-island effect using ASTER and ETM+ Data: Separation of anthropogenic heat discharge and natural heat radiation from sensible heat flux. *Remote Sensing of Environment*, 99(1), 44-54.
- Kanda, M., et al. "Roughness lengths for momentum and heat derived from outdoor urban scale models." *Journal of Applied Meteorology and Climatology* 46.7 (2007): 1067-1079.
- Kanda, Manabu, et al. "A new aerodynamic parametrization for real urban surfaces." *Boundary-layer meteorology* 148.2 (2013): 357-377.
- Kanda, M., Kawai, T., Kanega, M., Moriwaki, R., Narita, K., & Hagishima, A. (2005). A simple energy balance model for regular building arrays. *Boundary-Layer Meteorology*, 116(3), 423-443.
- Kanda, M., & Moriizumi, T. (2009). Momentum and heat transfer over urban-like surfaces. *Boundary-layer meteorology*, 131(3), 385-401.
- Kastner-Klein, Petra, and Mathias W. Rotach. "Mean flow and turbulence characteristics in an urban roughness sublayer." *Boundary-Layer Meteorology* 111.1 (2004): 55-84.
- Kawai, T., M. K. Ridwan, and M Kanda. "Evaluation of the simple urban energy balance model using selected data from 1-yr flux observations at two cities." *Journal of Applied Meteorology and Climatology* 48.4 (2009): 693-715.
- Kent, C. W., Grimmond, S., Barlow, J., Gatey, D., Kotthaus, S., Lindberg, F., & Halios, C. H. (2017). Evaluation of Urban Local-Scale Aerodynamic Parameters: Implications for the Vertical Profile of Wind Speed and for Source Areas. *Boundary-Layer Meteorology*, 164, 183–213 doi: [10.1007/s10546-017-0248-z](https://doi.org/10.1007/s10546-017-0248-z).
- Kormann, Robert, and Franz X. Meixner. "An analytical footprint model for non-neutral stratification." *Boundary-Layer Meteorology* 99.2 (2001): 207-224.
- Kotthaus, S., Smith, T. E., Wooster, M. J., & Grimmond, C. S. B. (2014). Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *ISPRS Journal Of Photogrammetry And Remote Sensing*, 94, 194-212.
- Kotthaus, Simone, and C. S. B. Grimmond. "Energy exchange in a dense urban environment–Part II: Impact of spatial heterogeneity of the surface." *Urban Climate* 10 (2014): 281-307.
- Lagouarde, J-P., et al. "Modelling daytime thermal infrared directional anisotropy over Toulouse city centre." *Remote Sensing of Environment* 114.1 (2010): 87-105.
- Lindberg, F., F. Olofson, T. Sun, C.S.B. Grimmond, C. Kent, W. Morrison. "Assessment and evaluation of ΔQs using different methods." UrbanFluxes Report D5.2 (2016)
- Lindberg, Fredrik, and C.S.B. Grimmond. "Nature of vegetation and building morphology characteristics across a city: influence on shadow patterns and mean radiant temperatures in London." *Urban Ecosystems* 14.4 (2011): 617-634.
- Lindberg F, CSB Grimmond, A Gabey, B Huang, CW Kent, T Sun, NE Theeuwes, L Järvi, H Ward, I Capel-Timms, YY Chang, P Jonsson, N Krave, DW Liu, D Meyer, KFG Olofson, JG Tan, D Wästberg, L Xue, Z Zhang 2018: Urban multiscale environmental predictor (UMEP) - An integrated tool for city-based climate services *Environmental Modelling and Software* 99, 70–87 [10.1016/j.envsoft.2017.09.020](https://doi.org/10.1016/j.envsoft.2017.09.020)
- London Grid for Learning, 2016. <https://weather.lgfl.org.uk/> . Accessed 24 July 2016

- Mahrt, L. "The bulk aerodynamic formulation over heterogeneous surfaces." Boundary-Layer Meteorology 25th Anniversary Volume, 1970–1995. Springer Netherlands, 1996. 87-119.
- Marconcini, Mattia, Wieke Heldens, Fabio Del Frate, Daniele Latini, Zina Mitranka, and Fredrik Lindberg. "EO-based products in support of urban heat fluxes estimation." In *Urban Remote Sensing Event (JURSE), 2017 Joint*, pp. 1-4. IEEE, 2017.
- Met Office, 2016 <http://www.metoffice.gov.uk/datapoint> .Accessed 24 July 2016
- Mitraka, Zina, Nektarios Chrysoulakis, Yiannis Kamarianakis, Panagiotis Partsinevelos, and Androniki Tsouchlaraki. "Improving the estimation of urban surface emissivity based on sub-pixel classification of high resolution satellite imagery." *Remote Sensing of Environment* 117 (2012): 125-134.
- Moriwaki, Ryo, and Manabu Kanda. "Scalar roughness parameters for a suburban area." *Journal of the Meteorological Society of Japan. Ser. II* 84.6 (2006): 1063-1071.
- Morrison W. S Kotthaus; S Grimmond; Atsushi Inagaki; Tiangang Yin; Jean-Philippe Gastellu-Etchegorry; Manabu Kanda; Christopher Merchant Longwave radiation fluxes observed with ground-based thermography to model urban thermal anisotropy, *Remote Sensing of Environment* (in review)
- Netatmo, 2016. [www.netatmo.com](http://www.netatmo.com) . Accessed 24 July 2016
- Owen, P. R., and W. R. Thomson. "Heat transfer across rough surfaces." *Journal of Fluid Mechanics* 15.03 (1963): 321-334.
- Ordnance Survey MasterMap, <https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html>, Accessed 11 August 2016
- Roth, Matthias. "Review of atmospheric turbulence over cities." *Quarterly Journal of the Royal Meteorological Society* 126.564 (2000): 941-990.
- Schmid, H. P. "Source areas for scalars and scalar fluxes." *Boundary-Layer Meteorology* 67.3 (1994): 293-318.
- Small, Christopher. "Multitemporal analysis of urban reflectance." *Remote Sensing of Environment* 81.2 (2002): 427-442.
- SPOT 2016, <http://www.intelligence-airbusds.com/> Accessed 25 July 2016
- Stull, Roland B. An introduction to boundary layer meteorology. Vol. 13. Springer Science & Business Media, 2012.
- Sugawara, Hirofumi. "Heat exchange between urban structures and the atmospheric boundary layer." *Doctoral Dissertation Tohoku Univ.* (2001).
- Sun, Jielun. "Diurnal variations of thermal roughness height over a grassland." *Boundary-Layer Meteorology* 92.3 (1999): 407-427.
- Theeuwes, N. E., Steeneveld, G. J., Ronda, R. J., Rotach, M. W., & Holtslag, A. A. (2015). Cool city mornings by urban heat. *Environmental Research Letters*, 10(11), 114022.
- USGS 2016 <http://landsat.usgs.gov/landsat8.php> Accessed 24 July 2016
- van der Loo, M., 2010. Extreme values, an R package for outlier detection in univariate data. *R package version*, 2.
- Van Ulden, A. P., and A. A. M. Holtslag. "Estimation of atmospheric boundary layer parameters for diffusion applications." *Journal of Climate and Applied Meteorology* 24.11 (1985): 1196-1207.
- Verma, S. B. "Aerodynamic resistances to transfers of heat, mass and momentum." *Estimation of areal evapotranspiration* 177 (1989): 13-20.
- Voogt, James A., and C. S. B. Grimmond. "Modeling surface sensible heat flux using surface radiative temperatures in a simple urban area." *Journal of Applied Meteorology* 39.10 (2000): 1679-1699.
- Voogt, James A., and Tim R. Oke. "Thermal remote sensing of urban climates." *Remote sensing of environment* 86.3 (2003): 370-384.
- Ward, H. C., J. G. Evans, and C. S. B. Grimmond. "Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in Swindon, UK." *Atmospheric Chemistry and Physics* 13.9 (2013): 4645-4666.
- Ward, Helen C. "Scintillometry in urban and complex environments: a review." *Measurement Science and Technology* 28.6 (2017): 064005.
- Weather Underground, 2016. [www.wunderground.com](http://www.wunderground.com) . Accessed 24 July 2016
- Weng, Q., Hu, X., Quattrochi, D. A., & Liu, H. (2014). Assessing intra-urban surface energy fluxes using remotely sensed ASTER imagery and routine meteorological data: A case study in Indianapolis, USA. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10), 4046-4057.
- Yang, K., Koike, T., Ishikawa, H., Kim, J., Li, X., Liu, H., Liu, S., Ma, Y. and Wang, J., 2008. Turbulent flux transfer over bare-soil surfaces: characteristics and parameterization. *Journal of Applied Meteorology and Climatology*, 47(1), pp.276-290.